Introduction to Statistical Population Modeling and Analysis for Pharmacokinetic Data

Marie Davidian

Department of Statistics, North Carolina State University

Raleigh, North Carolina 27695-8203, U.S.A.

davidian@stat.ncsu.edu

SUMMARY. The application of mathematical pharmacokinetic models in the area of risk assessment requires that variability and uncertainty be taken into appropriate account. We describe a standard statistical model framework in which these models may be placed that provides a basis for this and review methods for its implementation based on data.

KEY WORDS: Bayesian inference; Hierarchical model; Interindividual variability; Intraindividual variability; Markov chain Monte Carlo; Maximum likelihood; Nonlinear mixed effects.

Contents

1 Introduction

Pharmacokinetic (PK) models, especially physiologically based pharmacokinetic (PBPK) models, are mathematical models for mechanisms taking place within physiological systems that have emerged as important tools for human health risk assessment. These models have several potential uses, including serving as the basis for extrapolation from animals to humans

and from high to low doses and for improving the characterization of dose-adverse response relationships. The productive application of these models for these purposes demands that sources of information, including historical and experimental data, be exploited appropriately and requires that the variability and uncertainty involved in fitting these models to such data and then using them to develop extrapolations and risk estimates and to carry out other analyses be taken into adequate account. Specifically, policy recommendations based on these models must acknowledge the inherent variability in the populations for which the results of these analyses are intended and characterize faithfully the uncertainty associated with the analyses themselves.

These considerations thus require that mathematical PK and PBPK models be embedded in a statistical framework that provides the formal foundation for representing the relevant variability and uncertainty and hence for integrating the effects thereof into the risk assessment process when using these models. The broad goal of this document is to give an introduction to the basic statistical model framework into which PK models must be placed that acknowledges the sources of variability that must be considered and to show how the statistical model provides the springboard to methods for estimation of PK parameters and other quantities of interest in the model and to assessment of the inherent uncertainties involved in such estimation. We do not attempt to cover all relevant aspects of the application of PK models in the risk assessment context, nor do we provide a rigorous account of the formulation of PBPK models and their value in facilitating extrapolations and other analyses. Rather, we focus specifically on the basic statistical modeling and inferential considerations that must be appreciated.

We recognize that the backgrounds of readers may vary; thus, we have tried to achieve a balance between presenting faithfully some of the technical aspects where this is warranted

while making the conceptual and practical aspects accessible to readers without an extensive background in statistics. To assist readers with different backgrounds and expectations with navigation of the content at an appropriate level, the narrative provides frequent guidance on sections that may be skipped or skimmed by readers not wishing to delve into the some of the more technical aspects. Sections that are more technical are marked with an asterisk (*). Some readers may find it useful to skip these sections on an initial read and then approach them on subsequent review.

An outline of the organization of the material is as follows. In Section 2, we briefly review PK modeling in the traditional context of the study of pharmaceuticals and PBPK modeling in the setting of the study of potentially hazardous agents. Readers very familiar with such mechanistic modeling of physiological systems may wish to skip this section and proceed directly to Section 3. All readers may wish to review the PK/PBPK models presented in this section and in Figure 2, as they are used for illustration in subsequent sections.

To set the stage for introduction of the statistical framework, in Section 3 we discuss the notion of "population pharmacokinetic analysis." Readers well-versed in the principles and objectives of population analysis may wish to move directly to Section 4.

Section 4 describes the formal statistical model in which PK models must be embedded to carry out such "population" analyses, which involves a "hierarchy" that makes characterization of the sources of variability that must be considered transparent. This model is the essential element underlying these analyses, and all readers should cover at the least the less-technical subsections of Section 4 (not marked by asterisks).

Section 5 reviews methods that have been proposed historically and more recently to fit these models, i.e., estimate quantities of interest associated with them and provide accompanying estimates of uncertainty of estimation. An overview of the spirit of the methods is

provided at the outset of this section, followed by more technical accounts of the underpinnings of each. All readers should read the subsections of Section 5 without asterisks.

Finally, in Section 6, we discuss further analyses using PBPK models that may be based on and possible extensions of the framework without dwelling on technical considerations.

2 Pharmacokinetic Models

Readers familiar with pharmacokinetic modeling may wish to review models (1) and (3) on pages 7 and 10, respectively, and then proceed to Section 3.

Pharmacokinetics (PK) is the study of the time course and fate of the parent or metabolite concentrations of an agent (chemical, pharmaceutical, biologic) within a biological system (human, animal). Concentrations over time are determined by the rate and extent of the processes of absorption, distribution, metabolism, and excretion (ADME), where the term elimination usually refers to the combined effects of metabolism and excretion of the parent agent; Rowland and Tozer (1995) provide a comprehensive review of the basic concepts.

Of central interest is to characterize the ADME mechanisms that govern achieved concentrations over time. In the context of PK for pharmaceutical agents, understanding the nature of drug absorption, distribution, and elimination, both for specific individuals and across the population of individuals likely to receive the drug, is critical for development of effective dosing regimens that can achieve concentrations of a desired duration in a target range of therapeutic benefit. See Giltinan (2006) for an excellent review of considerations involved. For purposes of risk assessment, the goals are similar; understanding the ADME processes underlying concentrations of a potentially hazardous substance in different parts of the body, such as rates of its metabolism, for individuals and the population is important in its own right and for its relevance to further risk analyses, discussed shortly.

A standard approach to describing PK behavior is to represent the body (animal, human) by a series of *compartments*. Under a specified route of administration (of drug) or exposure (of hazardous agent), the compartmental representation leads to a system of differential equations in terms of *parameters* related to the ADME processes the describe mathematically the instantaneous rates of change of the amounts or concentrations of agent residing in each compartment based on assumptions on how the agent moves within and among the compartments. The solution of the system provides a formal mathematical description of the amounts or concentrations in the compartments at any time as a function of the parameters. A classic text on compartmental modeling is Gibaldi and Perrier (1982).

Ideally, the compartments in the model would represent real, identifiable components of the body. The models used in the study of PK of pharmaceutical agents, however, are typically gross simplifications of the true physiology and have thus been referred to as "empirical" models. Figure 1 shows a common model for the PK of an orally-administered drug, the one-compartment model with first-order absorption and elimination characteristics; here, the body is represented as a single "blood" compartment. At time $t = 0$, an oral dose D is introduced instantaneously into a hypothetical "absorption depot" (the gut) where the amount present is $A_a(t)$ at time t. Here, D represents the dose delivered into the system. Drug transfers into the "blood" compartment at fractional rate k_a and is eliminated at fractional rate k_e, where $A(t)$ is the amount of drug present in the compartment at time t, assumed to be perfectly mixed. Letting B be the bioavailability of the drug, the resulting linear differential equations are easily specified as

$$\begin{aligned} \frac{dA}{dt} &= k_a A_a - k_e A, & A(0) = 0, \\ \frac{dA_a}{dt} &= -k_a A_a, & A_a(0) = BD, \end{aligned} \tag{1}$$

and may be solved analytically for $A(t)$, leading to an expression for the concentration $C(t)$ at time t in the "blood" compartment given by

$$C(t) = \frac{A(t)}{V} = \frac{k_a DB}{V(k_a - k_e)}\{\exp(-k_e t) - \exp(-k_a t)\}, \quad k_e = Cl/V, \tag{2}$$

where V is the hypothetical volume required to account for all drug in the system; and Cl is drug clearance, the volume of drug cleared from the system per unit time. An expression for the concentration in the depot, $C_a(t)$, may also be derived. In analyses using (1)–(2) based on data on drug concentrations measured in the blood compartment only, B cannot be determined; here, it is standard to set it equal to a known constant determined from other information or to 1 in the absence thereof. Adopting this convention, from (2), concentrations of the drug in blood over time are dictated by the three *parameters* (k_a, Cl, V). We use the term "parameter" here to refer to quantities that govern a system but may be unknown (and whose values may be estimated from data); see Section 4.1 for more. These parameters all have meaningful interpretations, although they do not correspond precisely to specific physiological phenomena. It should be clear from (2) that, given the values of (k_a, Cl, V) for a specific individual and belief in the model, one can determine the concentration of drug at any time t following a dose of any magnitude D, even if the individual has never been administered this dose. It is this feature that enables pharmacokineticists to predict the time course of concentrations under different dosing regimens and thereby formulate regimens that achieve concentrations in a desired range.

It is important to recognize that (1)–(2) is a hypothesized mathematical description of the processes taking place within a *single* individual, and thus (k_a, Cl, V) govern the time course of concentrations that would be achieved for this individual only. Description of how the ADME processes represented by these parameters occur across the *population* is via a

statistical model, as discussed in subsequent sections.

Although models like (1)–(2) are grossly simplistic, the resulting descriptions of concentrations they yield and the basic information on ADME they embody nonetheless have been used quite successfully in drug development and evaluation. In the risk assessment arena, however, such simplified models do not facilitate many of the scientific objectives. Indeed, although the PK of pharmaceutical agents is relatively straightforward to study in the population of humans for which they are intended, PK studies of hazardous chemicals in humans are much more difficult to undertake, and, accordingly, studies in animals, such as rodents, are often more feasible, although human studies have been conducted (e.g. Bois et al., 1996; Gelman et al., 1996; Jonsson, Bois, and Johanson, 2001; Jonsson and Johanson, 2001; Mezzetti et al., 2003). If data are available for risk assessment only from animal experiments where the animals may have been subject to high exposure levels of the agent, their relevance to humans is predicated on the ability to make extrapolations, both from animal to human and from the high exposure levels in animal experiments to the lower levels likely to be encountered by humans, in regard to the concentrations of agent and its metabolites that may be achieved not only in the blood but in other parts of the body. In addition, in both animals and humans, the relationship between potentially toxic responses and the concentrations of toxic agent actually delivered to certain target tissues, often referred to as the "dose metric," rather than the administered dose or exposure, is likely to be a more valuable tool for understanding how the agent plays a role in adverse outcomes.

PBPK models have become an important tool for these problems; see Bailer and Dankovic (1997) for an excellent introduction. Relative to the simplistic PK models usually used for drugs, PBPK models attempt to provide a more "realistic" description of the body in terms of compartments representing identifiable physiological entities such as tissues or collections

of tissues and specific organs. The compartments and transfer among them are characterized by parameters denoting volumes, partition coefficients, blood flow and ventilation rates, metabolism rates, and so on, that lead to a formal mathematical system of equations. Figure 2 shows a schematic of a common PBPK model for an agent such as a volatile organic compound for which exposure is through the respiratory tract. The model involves four primary compartments representing well-perfused tissues; poorly-perfused tissues; fat tissues; and the liver, the site of metabolism of the agent, as well as the respiratory tract, which serves as an "exchange depot" compartment. In Figure 2, the parameters F_s are blood flow rates, $P_{s/\text{blood}}$ are tissue/blood partition coefficients, V_s are volumes, $s = \text{wp, pp, fat, liv}$ corresponding to the well-perfused, poorly-perfused, fat tissues and liver compartments; $V_{\max}$ and K_m are the maximum rate of metabolism and the Michaelis constant for the metabolism process in the liver; $P_{\text{blood/air}}$ is the blood/air partition coefficient; F_{alv} is the aveolar flow rate; and F_{card} is the total cardiac blood flow rate, where the ventilation-perfusion ratio is often defined as $VPR = F_{\text{alv}}/F_{\text{card}}$. The C_s are concentrations associated with the indicated compartmental sites indexed by s, including not only wp, pp, fat, and liv but also in inhaled and exhaled air and venous and arterial blood. Given a known exposure concentration (inhaled) and standard assumptions (e.g., flow-limited distribution, well-mixed compartments, instantaneous equilibrium between aveolar air, venous and arterial blood, etc.), the model implies that compartment-specific concentrations at time t are determined by determined by a system of equations

$$
\begin{aligned}
C_{\text{art}} = \frac{F_{\text{card}}C_{\text{ven}} + F_{\text{alv}}C_{\text{inh}}}{F_{\text{card}} + F_{\text{alv}}/P_{\text{blood/air}}}, \quad C_{\text{ven}} = \sum_s F_s C_s / F_{\text{card}}, \quad C_{\text{exh}} = (1-\delta)\frac{C_{\text{art}}}{P_{\text{blood/air}}} + \delta C_{\text{inh}} \\
\frac{dC_s}{dt} = \frac{F_s}{V_s}\left(C_{\text{art}} - \frac{C_s}{P_{s/\text{blood}}}\right), \quad s = \text{wp, pp, fat} \qquad (3) \\
\frac{dC_{\text{liv}}}{dt} = \frac{F_{\text{liv}}}{V_{\text{liv}}}\left(C_{\text{art}} - \frac{C_{\text{liv}}}{P_{\text{liv/blood}}}\right) - R_{\text{liv}} \ (s = \text{liv}), \quad R_{\text{liv}} = \frac{V_{\max}C_{\text{liv}}}{V_{\text{liv}}(K_m + C_{\text{liv}})},
\end{aligned}
$$

where δ is the assumed proportion of physiological dead space. Because of the Michaelis-Menten term for metabolic clearance in the liver, the system of differential equations in (3) is nonlinear. Hence, unlike the simple system (1), analytical solution of (3) to yield explicit expressions for the compartment-specific concentrations at time t is not possible, and for given parameter values must be carried out numerically using standard techniques for forward solution of differential equations (e.g., Gear, 1971).

In (3) and Figure 2, then, the compartment-specific concentrations of the agent present in the individual at any time t are governed by the flow rate (F, VPR), volume (V), partition coefficient (P), and metabolic ($V_{\max}$, K_m) parameters. Analogous to (1), given the values of these parameters corresponding to a particular individual, the concentration in any compartment at any time can be predicted for a given exposure setting/pattern to which that individual might be exposed. Again, we emphasize that model (3) is a hypothesized mathematical characterization for the PK processes within a single individual, so that the values of the parameters in (3) dictate concentrations for that individual only.

In the sequel, we use model (3) as a concrete basis with which to illustrate various issues associated with statistical modeling and analysis. However, the principles we demonstrate via this model apply in broad generality to any PBPK model, so should not be construed as relevant only to models similar to (3).

Identification of a suitable PK/PBPK model for a given application, e.g., choosing the number of compartments, must be based on understanding of the biological mechanisms thought to govern PK and the extent to which the model produces predicted concentration-time profiles similar to those actually observed on exposed individuals. A discussion of such PK model specification is beyond our scope here. In Sections 3–5, we assume that the analyst has a particular PK/PBPK model in mind, and turn to issues relevant to applying

that model to data and drawing inferences from it. Methods for doing so have been in place since the 1980s in the context of mostly empirical PK models and pharmaceuticals, and these same approaches are applicable to PBPK models. However, PBPK models are more complex and generally have a significantly larger number of parameters than do traditional empirical models, which introduces complications for statistical modeling and analysis, discussed later.

3 Population Pharmacokinetics

Readers familiar with the concepts and principles of population analysis may wish to proceed to Section 4.

Both simple empirical PK models like (1) and more complex PBPK models like (3) are deterministic mathematical models for the PK behavior of a single individual. One may be willing to believe that a particular such model is capable of describing the disposition of an agent for any individual in a population of interest. However, the values of most of the *PK parameters* in the model (e.g., flow rates, metabolic rates, compartmental volumes, etc.) for any specific individual, or indeed, for all individuals in the population, are not known in general. To address scientific objectives like those above, the values of the parameters or, more importantly, appropriate quantities related to them relevant to the objectives, discussed below, must be deduced indirectly from data. Before we may discuss estimation methods for this purpose, we must first clarify precisely the focus of such estimation efforts. The perspective of *population pharmacokinetic* modeling and analysis, first introduced in the context of the study of pharmaceutical agents (Sheiner, Rosenberg, and Melmon, 1972; Sheiner, Rosenberg, and Marathe, 1977; Beal and Sheiner, 1982; Sheiner and Ludden, 1992) provides this clarification and is equally relevant to the implementation of PBPK models for risk assessment (Bois et al., 1996; Jonsson and Johanson, 2001ab; Jonsson et al., 2001).

Population PK modeling and analysis takes the view that scientific interest focuses primarily on how mechanisms underlying PK behavior of individuals take place overall in the *population* (rather than in any one specific individual). More precisely, as these mechanisms are represented explicitly for each individual in the population by his/her parameters in the chosen PK model, this amounts to interest in how the values of the parameters occur across the entire population and in particular how they *vary* across the population. In the context of pharmaceuticals, this perspective arises in part from the need to set dosing recommendations for a population. Because an individual's parameters dictate achieved drug concentrations that are assumed associated with therapeutic effect, variability in them across individuals may translate into variability in concentrations and hence in effect. If this variability is large, devising broadly applicable recommendations that yield satisfactory therapeutic benefit to the majority of individuals may be difficult. Hence, understanding this variability is critical. Likewise, in a risk assessment context, setting policy requires an understanding of the extent to which individuals in the population vary in metabolizing toxic compounds and in the concentrations they may achieve in target tissues, which are in turn related to adverse response; we discuss this further in Section 6. One may also be interested in specific individuals, and this is also accommodated by the population perspective; however, the view is that the broader scientific questions relevant to policymaking naturally involve understanding of variability in the population and hence require consideration of the population as a whole.

It is well-established that individuals from diverse populations such as human populations can vary considerably in their underlying ADME mechanisms and thus in the values of the parameters in the model. Figure 3, from a human PK study of the drug theophylline given orally in the same dose to all subjects, provides a simple illustration: although the drug concentration-time profiles exhibit a shape consistent with (2) for each subject, suggesting

that this model provides a good description for all, the specific form varies from subject to subject, which may be attributed to variability in the values of (k_a, Cl, V) across subjects.

Thus, population PK modeling and analysis involves specifying a suitable (statistical) framework in which variability in PK model parameters across the population of interest may be represented and estimated based on observable data from multiple individuals. As we formalize in the next section, the focus of estimation is thus on quantities that describe the *distribution* of PK parameters across the population; these quantities are referred to as *population parameters*. For example, the mean and variance of rate of metabolism $V_{\max}$ in the population are examples of such population parameters, where the variance is the population parameter explicitly quantifying how $V_{\max}$ values vary across individuals, and the mean quantifies the average or "typical" value of $V_{\max}$ in the population about which values vary. PK parameters corresponding to specific individuals may also be estimated; however, it is the population parameters that are most relevant to the scientific objectives.

The population approach may be further refined. Some of the population variability may in fact be attributed to systematic associations among parameter values and known characteristics of individuals, such as gender, ethnicity, health status, genotypic information, etc. We cite a specific example in the context of a toxicokinetic study in Section 4.6. In population PK studies of drugs, where the number of individuals may be large (hundreds), population analysis may involve the additional step of trying to estimate quantities that describe such associations (e.g., Maitre et al., 1992; Mandema et al., 1992). If some of the population variability in parameters (and hence achieved concentrations) can be linked to known attributes, targeted dosing recommendations taking these features into account may be developed, improving outcomes. Ideally, a similar perspective would be critical for risk assessment; however, studies involving sufficient numbers of human subjects to identify real-

istically true such associations are more problematic to carry out with hazardous agents; see Mezzetti et al. (2003) for an example of a toxicokinetic study involving over 100 subjects for which this was attempted in a population analysis. In our formal description of the statistical framework for population analysis in the next section, we thus mention the extension to this case briefly.

4 Hierarchical Statistical Model

Intuitively, learning about how PK parameter values vary across individuals and more generally how they are distributed in the population requires concentration data over time collected from each of a sample of individuals from the population. Thus, population analysis is usually predicated on the availability of such data, and the statistical framework is described in terms of them. We now provide a formal, careful description of these data and the basic model framework in the context of a single, typical experimental PBPK study. As we discuss later, even if data of the form described below are not available, the model still provides a conceptual framework for thinking about how data in other forms might arise.

All readers should read Sections 4.1, 4.2, 4.4, 4.6, 4.5, and 4.7. Readers desiring a more technical description of some aspects of the statistical model should also read Section 4.3.

Description of the model and its subsequent use requires definition of a good deal of notation. For convenience, Table 1 provides a summary of the various symbols defined in the sequel and their interpretation.

4.1 *Basic Set-Up and Model*

Suppose the study involves N individuals drawn from a population of interest, indexed by i. In Section 6, we discuss extensions of the basic model to more complicated data structures.

Each individual i is exposed to the study agent at a known level by a known route of exposure. For example, each individual might be placed in an inhalation chamber in which the concentration of agent is set at a particular level (10 ppm benzene, say) for a known duration (4 hours, say), so that C_{inh} in (3) is (ideally) equal to the exposure level during the time spent in the chamber and zero after exit from the chamber. We denote all information on the exposure level and duration undergone by individual i as $\boldsymbol{E}_i$. Suppose that the analyst has in mind a particular PBPK model to be applied to the data.

On each individual i, concentration measurements may be taken on what correspond to one or more of the compartment-specific concentrations C involved in the PBPK model at n_i known time points, denoted as t_{ij}, where $j = 1, \ldots, n_i$ indexes the time points, and time may be scaled so that time 0 represents a milestone such as the beginning of the exposure or measurement period. For example, concentrations in exhaled air and venous blood [C_{exh}, C_{ven} in (3)] might be measured on each individual at a few times during exposure and then over the next few days or week, so that $t_{i1}, t_{i2}, \ldots, t_{in_i}$ correspond to the n_i ordered measurement times for individual i, with time 0 representing time of exposure initiation. Assume c compartment-specific concentrations are sampled on all individuals (so $c = 2$ in our example above), where $c \leq$ the total number of compartment-specific concentrations involved in the PBPK model. Let Y_{ijk} be the kth compartment-specific concentration measured among the total of c measured in the study, $k = 1, \ldots, c$, on individual i at time t_{ij}. Collecting all c concentration measurements at time t_{ij} into a vector $\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \ldots, Y_{ijc})'$ of size $(c \times 1)$, we may summarize all concentration measurements taken on individual i over time in the study succinctly as $\boldsymbol{Y}_i = (\boldsymbol{Y}'_{i1}, \ldots, \boldsymbol{Y}'_{in_i})'$, which is of length cn_i. The numbers and values of the time points may be different for different individuals and for different compartment-specific concentrations k on the same individual; see Section 6. We consider the case of "common"

measurement times for all c concentrations for simplicity, so that the presentation is parallel to that in the traditional population PK literature.

In addition to the concentration measurements, physiological measurements may be recorded on each individual, e.g., body weight, height, blood/air partition coefficient, fat free body mass, and so on; although these are often referred to as "physiological parameters," we call them *measurements* to distinguish them from other PK parameters and population parameters defined shortly, so that, following statistical modeling convention, we reserve use of the term "parameter" to denote an unknown quantity that might be estimated from data. For individual i, we denote the collection of these as $\boldsymbol{\phi}_i$. Some of these may be used to calculate numerical values for some parameters in the PBPK model for each individual, e.g., $V_{\text{fat},i}$, the value of V_{fat} in (3) for individual i, which may then be taken as known. They may also be used to scale some of the other PK parameters that are potentially unknown, so as to take into account known physiological dependencies between them and these PK parameters. For example, volumes such as V_{wp} may be reexpressed in (3) as fractions of lean body weight, and other parameters may be expressed in terms of physiological measurements and other (unknown) model parameters. There may also be a need to transform some parameters so that they obey known constraints; e.g., fractional blood flow rates must sum to 1. We denote the potentially unknown PK parameters for individual i in the PBPK model, possibly rescaled and transformed, as $\boldsymbol{\theta}_i$, and denote its length by p. Thus, for example, $V_{\text{max},i}$, an element of $\boldsymbol{\theta}_i$, denotes the maximum rate of metabolism for individual i.

Additional attributes of each individual may also be collected; e.g., gender, ethnicity, genotype, and so on, that may not be straightforwardly incorporated in the PBPK model. Denote the collection of these attributes as $\boldsymbol{A}_i$ for individual i. The $\boldsymbol{A}_i$ do not enter into the basic model; we discuss an extension of the model including them in Section 4.6.

With these definitions, we may summarize the information available in the study as $(\boldsymbol{Y}_i, \boldsymbol{E}_i, \boldsymbol{A}_i, \boldsymbol{\phi}_i)$, $i = 1, \ldots, N$. The $\boldsymbol{E}_i$ are fixed by the design of the study. On the other hand, from a statistical perspective, $(\boldsymbol{Y}_i, \boldsymbol{A}_i, \boldsymbol{\phi}_i)$ may be viewed as *random vectors* representing the data that would be collected in such a study, which take on specific numerical values once the study has been conducted. A standard assumption made in almost all PBPK analyses based on the hierarchical model below is that the sets of random vectors $(\boldsymbol{Y}_i, \boldsymbol{A}_i, \boldsymbol{\phi}_i)$ are *statistically independent* across i, which would be the case if the individuals are reasonably thought to be unrelated, so that, e.g., the way the concentrations might turn out for one individual is unrelated to the way they might turn out for another. The $\boldsymbol{\theta}_i$, $i = 1, \ldots, N$, are also viewed as independent random vectors, where the independence implies that the PK processes taking place inside any individual do not have a bearing on those for any other. However, most of the elements of $\boldsymbol{\theta}_i$, such as metabolic parameters, are not observed. Recalling Section 3, estimating population parameters that describe how the $\boldsymbol{\theta}_i$ are distributed in the population is of central interest in a population analysis.

We may now state the basic form of a *hierarchical statistical model* formalizing our beliefs regarding how the observed data from such a study may arise. From the discussion following (3), the solution to the system of equations defining a PBPK model, which gives expressions for the compartment-specific concentrations in the model at any time t, usually must be obtained numerically. Nonetheless, we may conceptualize for the kth of the c compartment-specific concentrations collected that the solution is a function f_k, $k = 1, \ldots, c$, of time t, exposure level, observed physiological measurements, and unobserved PK parameters. Thus, for individual i, the PBPK model in principle yields an expression (that may be calculable only numerically) for the kth compartment-specific concentration, $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$, at any

time t, including at the observation times t_{ij}. We may collect these into a vector

$$\boldsymbol{f}(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) = \{f_1(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i), \ldots, f_c(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)\}' \tag{4}$$

that summarizes the expressions for all c compartment-specific concentrations sampled in the study for individual i at any time t.

Because the $\boldsymbol{Y}_i$ are random vectors, from a statistical perspective, the Y_{ijk} are random variables that, given the values of $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ for individual i, take on their values according to some *probability distribution*; we discuss this further in Sections 4.2 and 4.3. In population PBPK analyses, the Y_{ijk} are usually assumed to follow a lognormal or normal probability distribution. When the magnitude of intraindividual variability is is not too large, as is typically the case for individual PBPK data, the lognormal and normal distributions are almost indistinguishable. Thus, for simplicity in describing how the model incorporates a description of intraindividual variability, we initially write the part of the model describing individual behavior in a form that is suitable to assuming Y_{ijk} are normally distributed, as is standard in the pharmaceutical population PK literature (e.g., Beal and Sheiner, 1982). However, the considerations we discuss in Sections 4.2 and 4.3 are equally relevant when assuming lognormality or yet other models; we discuss this further below.

With these considerations, we write the model as a two-stage hierarchy:

$$\textit{Stage 1: Individual-Level Model} \qquad \boldsymbol{Y}_{ij} = \boldsymbol{f}(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) + \boldsymbol{e}_{ij}, \;\; j = 1, \ldots, n_i, \tag{5}$$

$$\textit{Stage 2: Population Model} \qquad \log(\boldsymbol{\theta}_i) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \;\; i = 1, \ldots, N, \tag{6}$$

where "$\sim$" denotes "distributed as" and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and *covariance matrix* $\boldsymbol{\Sigma}$.

The "stage 1" *individual-level model* (5) describes how concentrations observed *on individual* i are thought to arise. Thus, (5) pertains *only* to individual i and would be the statistical

model used to describe i's concentrations if i were the only individual of interest. Here, $\boldsymbol{f}$ is as defined in (4), so represents the collection of expressions for the c compartment-specific concentrations dictated by (deterministic) PBPK model at each time. It is well-appreciated that actual, measured concentrations do not correspond exactly to concentrations determined by the model; thus, the elements of the "deviation" random vector $\boldsymbol{e}_{ij} = (e_{ij1}, e_{ij2}, \ldots, e_{ijc})'$ represent the amounts by which measured concentrations Y_{ijk} in $\boldsymbol{Y}_{ij}$ deviate from those dictated by the model. Zeroing in on the the kth of the c measured compartment-specific concentrations, Y_{ijk}, (5) implies

$$Y_{ijk} = f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) + e_{ijk}, \quad k = 1, \ldots, c. \tag{7}$$

In (5) and (7), the e_{ijk} are also random variables. (7) has the form of a "nonlinear regression model" with unknown "regression parameters" $\boldsymbol{\theta}_i$ (e.g., Davidian and Giltinan, 1995, Ch. 2), for which it is standard to assume that the "deviations" are normally distributed with mean 0, so that the Y_{ijk} are also normal with mean $f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$, viewing $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ as fixed quantities. An alternative formulation approximately equivalent to assuming that the Y_{ijk} given $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ are lognormally distributed with mean $f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ is

$$\begin{aligned} \log(Y_{ijk}) &= \log\{f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)\} + e_{ijk}, \quad k = 1, \ldots, c, \\ \log(\boldsymbol{Y}_{ij}) &= \log\{\boldsymbol{f}(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)\} + \boldsymbol{e}_{ij} \end{aligned} \tag{8}$$

In (8), although we use the same symbol for these "deviations" e_{ijk}, they are on a different (log) scale than those in (7). In Section 4.2, we discuss the phenomena that may be responsible for deviations in either model, which have to do with *intraindividual variability*.

The "stage 2" *population model* is given this name because it operationalizes the main goal of population analysis; i.e., it describes how the unobserved PK parameters $\boldsymbol{\theta}_i$ are distributed in the population of individuals by formally representing how the $\boldsymbol{\theta}_i$ take on their

values in the population by a *probability distribution*. It is well-recognized that PK parameters, like many biological quantities, tend to have skewed distributions in the population. Accordingly, it is standard to assume that each component of the random vectors $\boldsymbol{\theta}_i$ takes on its values in the population according to a lognormal distribution. This is usually represented equivalently in population PBPK analyses by assuming that the random vector containing the logarithms of the elements of $\boldsymbol{\theta}_i$, which we write as $\log(\boldsymbol{\theta}_i)$, takes on values according to a multivariate normal distribution, as in (6). Thus, if $\boldsymbol{\theta}_i$ is p-dimensional, this is a p-variate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)'$, and each of the elements μ_ℓ corresponds to the mean of one of the PK parameters in $\boldsymbol{\theta}_i$ on the log scale; e.g., the mean of $\log(V_{\max,i})$ values across the population of all possible individuals. The mean vector $\boldsymbol{\mu}$ and its elements are thus *population parameters* in the sense described in Section 3. The covariance matrix $\boldsymbol{\Sigma}$ is also a population parameter; its elements characterize aspects of the variability of the (log) PK parameters in the population. The diagonal elements of the covariance matrix $\boldsymbol{\Sigma}$ are the variances of the logarithms of the p PK parameters in $\boldsymbol{\theta}_i$, which we write as $(\Sigma_1^2, \Sigma_2^2, \ldots, \Sigma_p^2)$, where Σ_ℓ^2 is the variance in the population of the ℓth element of $\log(\boldsymbol{\theta}_i)$. Because of the relationship between the normal and lognormal distributions, the square roots of these variances, the standard deviations Σ_ℓ, are approximately equal to the *coefficients of variation* (CV) of the PK parameters on their original scales, which is an alternative, probably more meaningful measure of variability for skewed populations.

A further aspect of variability in the population is represented $\boldsymbol{\Sigma}$, that of *covariability*. The off-diagonal elements of $\boldsymbol{\Sigma}$ are the *covariances* between pairs of log PK parameters in the population, which may be transformed to the *correlation* scale. The covariances, and equivalently correlations, quantify associations (on the log scale) among the PK parameters in the population; e.g., if individuals with high maximum metabolic rate $V_{\max,i}$ also tend to

have high flow rate $F_{\text{liv},i}$, these quantities would be positively correlated in the population. Knowledge of such associations may be useful interpreting why certain subpopulations of individuals exhibit certain dose-response relationships, for example.

We discuss the population model further in Section 4.6.

Summarizing, the hierarchical statistical model (5)–(6) describes how concentration-time data in a PBPK study of multiple individuals arise. The second stage population model characterizes variability in PK in the population by representing individuals in the population by their PK parameters, which are assumed to take on values in the population according to the lognormal probability distribution. Thus, we may view the individuals in a study as being represented by "random draws" from this probability distribution through their $\boldsymbol{\theta}_i$. At the first stage, the process by which individuals represented by their specific $\boldsymbol{\theta}_i$ give rise to measured concentrations is characterized. A notable feature of this individual-level model is that it *embeds* the deterministic, mathematical PBPK model into a statistical framework. This statistical framework says that, viewing the $\boldsymbol{\theta}_i$ for a specific individual as a fixed quantity that determines his/her concentrations, measured concentrations follow the trajectory dictated by the PBPK model, but deviate from it due to sources of variability operating *within* individuals, whose effects are represented by the $\boldsymbol{e}_{ij}$ in (5), discussed next.

Statistical models like (5)–(6) are also referred to as *nonlinear mixed effects models* or *hierarchical nonlinear models* in the statistical literature.

4.2 *Sources of Intraindividual Variability*

To complete the specification of the stage 1 individual-level model, the analyst must consider carefully the phenomena that are thought to lead to the fact that observed, measured concentrations do not tend to coincide with those dictated by the PBPK model, but rather

exhibit variability about the deterministic concentration trajectories it determines. It is critical that the sources of this *intraindividual* variability be taken into adequate account. Why? There is also *interindividual* variability (among $\boldsymbol{\theta}_i$) in the population. These two sources of variability combine to produce the overall pattern of *total variability* in data that might arise from a PBPK study. The hierarchical model (5)-(6) effectively *partitions* this total variability into its intra- and interindividual components. Thus, because the total variability stays the same, misrepresentation of the variability from one source leads to an inaccurate partition of the variability, and thereby can lead to misrepresentation from the other. Thus, failing to characterize faithfully the nature of intraindividual variability could compromise a key focus of population analysis, that of quantifying interindividual variability in the population.

Unfortunately, the considerations involved in thinking about intraindividual variability are generally not emphasized sufficiently in the literature, and many practitioners have a misimpression that variability of observed concentrations about the PK model trajectory are due entirely to "measurement error." This may be an unfortunate consequence of the tendency in the statistical literature to refer to deviations like $\boldsymbol{e}_{ij}$ and e_{ijk} in (5), (7), and (8) as "errors." While errors in measurement are certainly one key component of intraindividual variability, there are others whose relative magnitude may be nonnegligible. We now give a conceptual discussion of some of these sources of variation, which all readers should review. Section 4.3 provides a technical description of how assumptions on these sources of variation may be formalized in writing down the full statistical model and may be skipped by readers uninterested in these details.

Consider a single compartment-specific concentration in the PBPK model; e.g., C_{exh} in (3), labeled as the kth among the c compartment-specific concentrations measured in the study. Focusing on a specific individual i, suppose the observed, measured C_{exh}, Y_{ijk}, for i

are postulated to follow the individual-level model (7). Here, because we are "zeroing in" on individual i, we regard i's exposure pattern $\boldsymbol{E}_i$, physiological measurements $\boldsymbol{\phi}_i$, and PK parameters $\boldsymbol{\theta}_i$ as fixed quantities determining i's data. Thus, in the statistical model (7), the deviations e_{ijk} at times t_{ij} represent how the measured concentrations Y_{ijk} end up scattered about the expressions $f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i\boldsymbol{\theta}_i)$ dictated by the PBPK model and i's particular $\boldsymbol{E}_i$, $\boldsymbol{\phi}_i$, and $\boldsymbol{\theta}_i$ values.

Figure 4 presents a conceptual perspective on two possible sources on intraindividual variability that may be responsible for this scatter. The solid black line represents the deterministic, continuous time trajectory $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ dictated by the PBPK model at all times t under the exposure experienced by i ($\boldsymbol{E}_i$) and the physiology and PK dictated by i's values $\boldsymbol{\phi}_i$ and $\boldsymbol{\theta}_i$. Realistically, the trajectory traced by f_k may not capture all within-individual physiological processes perfectly; e.g., in an inhalation study, an individual's breathing pattern over time may exhibit natural changes, while in a study sampling urine concentrations there may be variation in the way in which excretion occurs over collection times. We may thus conceptualize that the PBPK model, which yields "smooth" a concentration trajectory, can almost, but not quite, capture the *true* concentration profile *actually realized* over all t, which is influenced by these phenomena. This true profile, if it could be seen exactly for all t, might look like the solid gray line, which "tracks" the solid line but shows some departures from (variability about) it for these reasons.

Now, in the study, concentrations are only sampled at the intermittent times t_{ij} for individual i. If the true, actually realized concentrations could be ascertained perfectly, the recorded values for them would be those at each t_{ij} on the solid gray line. However, as an assay must be used to quantify the concentrations in samples taken at each t_{ij}, assay measurement error may be introduced, so that the measured values actually recorded at

each t_{ij} correspond to the solid diamond symbols in the figure. The deviation e_{ijk} in the first stage statistical model (7) thus represents the combined effect of actual realized concentration (solid gray line) and measurement error (diamonds) yielding observed concentrations that deviate from the solid black deterministic trajectory traced by the model, $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$.

Now consider that the true, gray concentration profile that i ends up having in the study (we cannot observe it but can conceptualize its existence as above) is only one of many such realized true profiles that individual i could have produced under the circumstances of the study. We can imagine that if the same study of i were repeated over and over, the realized true profiles each time might be different. For example, i's breathing pattern might be slightly different during each possible study of i we might do. Similarly, the assay errors contaminating the particular true concentrations on i that occurred in the study are representative of errors that could be committed when quantifying concentrations in such samples, so that if the same study on i was repeated, the errors committed would be different every time because of natural variability in the performance of the assay.

With this in mind, we can now interpret the model (7) a little more precisely. As noted on page 19, the e_{ijk} are assumed statistically to have "mean 0." This means that if we could average over all possible realized true profiles that i could have and all possible assay errors that could be committed if individual i were observed under the conditions of the study, the effects of "realization variability" and "measurement error variability" would all "average out" to zero. The result is that we may interpret $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ given by the deterministic PBPK model as the "inherent trajectory" that specifies how i's concentrations would tend to evolve over time "on average" under the conditions of the study, where the "average" is over all possible realizations of actual true concentration profiles that i might have and measurement errors that could be committed. The PK parameters $\boldsymbol{\theta}_i$ may thus be viewed

as an "inherent characteristic" of individual i dictating this tendency. Hence, a fundamental principle underlying population analysis under the hierarchical statistical model (5)-(6) is that such "inherent" properties of individuals are of central scientific interest.

Of course, once we conduct the study, the data values we actually see are the combined result of one possible realized true concentration profile and one set of measurement errors.

Our conceptual depiction identifies two potential sources of intraindividual variability: "realization" variability due to the tendency for true concentrations to deviate from the "smooth" behavior dictated by f_k, and "measurement error" variability. This suggests that we might rewrite (7) as

$$Y_{ijk} = f_k(t_i j, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) + e_{R,ijk} + e_{M,ijk}, \quad e_{ijk} = e_{R,ijk} + e_{M,ijk}, \tag{9}$$

where $e_{R,ijk}$ represents the part of the overall deviation e_{ijk} from f_k due to "realization" variability and $e_{M,ijk}$ that due to "measurement error" variability. The analyst's job in completing the specification of (5) boils down to making realistic assumptions on the $e_{R,ijk}$ and $e_{M,ijk}$; in particular, the probability distributions that describe how they take on their possible values across all possible realizations and measurement errors. For example, it may be assumed that "realization deviations" $e_{R,ijk}$ take on their values according to a $\mathcal{N}(0, \sigma^2_{R,k})$ distribution, where $\sigma^2_{R,k}$ is a variance characterizing the magnitude of variability of "realized profiles" about the deterministic trajectory $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$, where the "$k$" subscript on $\sigma^2_{R,k}$ indicates that this variability pertains to the kth compartment-specific concentration. Likewise, the "measurement error deviations" $e_{M,ijk}$ may be assumed to arise following a $\mathcal{N}(0, \sigma^2_{M,k})$ distribution, where $\sigma^2_{M,k}$ is a variance quantifying the extent of measurement (assay) error, where again the "k" subscript reminds us that this measurement error variance pertains to the assay used to quantify the kth compartment-specific concentration. It is often

assumed that these two sources operate independently, so that the "overall" intraindividual variance is

$$\sigma_k^2 = \sigma_{R,k}^2 + \sigma_{M,k}^2. \tag{10}$$

The analyst must moreover acknowledge that the foregoing considerations apply to *each* of the $k = 1, \ldots, c$ compartment-specific concentrations that may be measured on each individual in a study. For example, there would be an "overall" intraindividual variance σ_k^2 associated with each of the $k = 1, \ldots, c$ measured compartment-specific concentrations.

There are yet further issues that must be considered. First, for a particular compartment-specific concentration k, from inspection of Figure 4, note that "realized" concentration values on the gray line close together in time tend to occur "on the same side" of the "inherent trajectory" $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ (solid line). Statistically speaking, this suggests that we might expect the $e_{R,ijk}$ in (9), and hence concentration measurements Y_{ijk} for the same k to be *correlated* over time, where the correlation is stronger the closer together in time two measurements are taken. In fact, a bit of thought reveals that we might expect the form of the "realized" profiles for two different compartment-specific concentrations k and k', say, to be associated somehow. For example, if the "realized" (gray) profile for compartment k is "high" relative to the "on average" trajectory $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ at a particular time because of how the breathing pattern is manifesting at this time, physiologically, the "realized" profile for compartment k' might be correspondingly "low" relative to its "on average" trajectory $f_{k'}(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ at the same time. Again, this would lead to *correlation*, this time between measurements Y_{ijk} and $Y_{ijk'}$ taken at the same time point.

One can envision other associations among the way deviations occur that might lead to still other correlations among measured compartment-specific concentrations. We do not

discuss this further, but we remark that modeling intraindividual variability appropriately requires that all suspected correlations be acknowledged in the statistical model. Failure to take into account strong such "intraindividual" correlations could lead to a serious misrepresentation of the overall pattern of intraindividual variability, and in turn to misrepresentation of interindividual variability, as discussed at the beginning of this section. Almost all published population PBPK analyses ignore intraindividual correlation entirely. The extent to which this compromises PBPK population analyses is not known and deserves careful study.

Section 4.3 offers a more technical formulation that formalizes these issues. Readers not interested in these details may wish to proceed to Section 4.4, which summarizes the full individual-level model that, rightly or wrongly, is usually assumed in published PBPK population analyses.

4.3 *Intraindividual Variability Modeling Considerations**

We now give a more technical treatment of intraindividual variability modeling along the lines of that in Davidian and Giltinan (2003, sec. 2.2.2) for readers interested in the details.

Consider again as in Section 4.2 the kth compartment-specific concentration measured in the study and the individual-level model (7). We may now be more precise about what this model assumes. As noted previously, we are focusing here exclusively on specific individual i; thus, we regard i's exposure pattern $\boldsymbol{E}_i$, physiological measurements $\boldsymbol{\phi}_i$, and PK parameters $\boldsymbol{\theta}_i$ as fixed quantities determining i's data. Thus, technically, when we say that the e_{ijk} in (7) are taken to have "mean 0," we mean that, *conditional* on these quantities, the "deviations" e_{ijk} have mean 0, which is written as $E(e_{ijk}|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) = 0$. This implies that (7) specifies that the mean of measured concentrations Y_{ijk} at t_{ij}, *conditional* on $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$, is

$$E(Y_{ijk} \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) = f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i). \tag{11}$$

Accordingly, when we refer in Section 4.2 to $f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ as describing the "inherent trajectory" specifying how i's concentrations would arise over time "on average," we mean formally that $f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ is the *conditional mean* (11). Similarly, in (9), the "realization" deviations $e_{R,ijk}$ and the "measurement error "deviations" $e_{M,ijk}$ are assumed to have conditional means equal to zero; i.e., $E(e_{R,ijk}|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) = 0$ and $E(e_{M,ijk}|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) = 0$ so that $E(e_{ijk}|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i) = 0$. We now consider more formally the considerations involved in specifying assumptions on what are really the *conditional probability distributions* of these deviations and hence of the Y_{ijk}; that is, given the particular values $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$.

First, consider $e_{R,ijk}$. The variance of the conditional probability distribution of the $e_{R,ijk}$ given $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ quantifies the variability these deviations exhibit about the "inherent trajectory" $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ in producing a realized gray curve like that in Figure 4. As in the previous section, we might assume that this variance is the same at all times and equal to some value $\sigma^2_{R,k}$. If we further believe that the probability distribution of Y_{ijk} values given $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ is normal, as in Section 4.1, then it would be natural to assume that the conditional probability distribution of the $e_{R,ijk}$ at any t_{ij} is $\mathcal{N}(0, \sigma^2_{R,k})$. In addition, as we noted in Section 4.2, from Figure 4, two true realized concentrations on the gray line at times close together tend to occur "on the same side" of the "inherent trajectory." This implies that for two observation times t_{ij} and $t_{ij'}$, say, sufficiently close in time, $e_{R,ijk}$ and $e_{R,ij'k}$ would tend to be positive or negative together. On the other hand, realized values at times far apart bear little relation to each other. Formally, this suggests that the "realization deviations" are likely to be positively correlated within an individual, with the strength of the correlation "damping out" as the time points become more separated. We thus consider the conditional probability distribution of all the $e_{R,ijk}$, $j = 1, \ldots, n_i$, simultaneously. Placing these in a vector $\boldsymbol{e}_{R,ik} = (e_{R,i1k}, e_{R,i2k}, \ldots, e_{R,in_ik})'$, under the assumption that each $e_{R,ijk}$

has a normal distribution as above, the obvious assumption is that

$$\boldsymbol{e}_{R,ik} \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{V}_{R,ik}), \quad \boldsymbol{V}_{R,ik} \ (n_i \times n_i), \tag{12}$$

where this notation reminds us that we are treating $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ as fixed quantities, and $\boldsymbol{V}_{r,ik}$ is a covariance matrix; i.e., $\boldsymbol{e}_{R,ik}$ has a n_i-variate normal distribution. The diagonal elements of $\boldsymbol{V}_{R,ik}$ are all equal to $\sigma^2_{R,k}$. The off-diagonal elements correspond to the covariances, and hence the correlations, among pairs of the $e_{R,ijk}$. Expressions for these off-diagonal elements that would be candidates for describing the kind of "damped," serial correlation anticipated over time are described by Diggle et al. (2001, Ch. 5); one popular model takes the correlation between two deviations $e_{R,ijk}$ and $e_{R,ij'k}$ at times t_{ij} and $t_{ij'}$ to be of the form $\exp(-\alpha|t_{ij} - t_{ij'}|)$, where α is a constant to be estimated.

The $e_{M,ijk}$ represent the deviations associated with measuring true realized concentrations using an error-prone assay. Valid measuring techniques and devices tend to commit haphazard errors over repeated uses; thus, there is no reason to believe that $e_{M,ijk}$ and $e_{M,ij'k}$ associated with measured concentrations at two times t_{ij} and $t_{ij'}$ are correlated, no matter how close or far apart the times are. The assumption about the variance of $e_{M,ijk}$ should follow from knowledge of the assay. Some assay procedures tend to commit errors of the same magnitude no matter what the level in the sample, in which case taking the variance to be a constant $\sigma^2_{M,k}$, as in Section 4.2, at all times is reasonable. Other assays tend to commit multiplicative errors. This is often handled in PK analysis by working on the log scale; i.e., using model (8) rather than (7), as on the log scale the errors tend to be of the same magnitude, making the assumption of a constant variance reasonable. Indeed, all of the foregoing considerations on intraindividual variability may be applied equally to the model (8) if a lognormal conditional distribution for Y_{ijk} is posited. For definiteness, continuing

with our description in terms of (7), assuming constant measurement error variance, letting $\boldsymbol{e}_{M,ik} = (e_{M,i1k}, e_{M,i2k}, \ldots, e_{M,in_ik})'$, and taking each $e_{M,ijk}$ to be conditionally normally distributed, analogous to (12), we have

$$\boldsymbol{e}_{M,ik} \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_{M,ik}), \quad \boldsymbol{V}_{M,ik} \ (n_i \times n_i), \tag{13}$$

where, owing to the assumed lack of correlation, $\boldsymbol{V}_{M,ik}$ is a diagonal covariance matrix with all diagonal elements equal to $\sigma^2_{M,k}$.

Combining (12) and (13) leads to an assumption for the probability distribution of the overall deviations e_{ijk} given $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$. If we are willing to believe that that the measurement process produces measurement error deviations like those in (13) regardless of the nature of the true realized profile, then we may assume that the random vectors $\boldsymbol{e}_{R,ik}$ and $\boldsymbol{e}_{M,ik}$ are statistically independent, which leads to the assumption on the e_{ijk} given by

$$\boldsymbol{e}_{ik} \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_i), \quad \boldsymbol{V}_{i,k} = \boldsymbol{V}_{R,ik} + \boldsymbol{V}_{M,ik}. \tag{14}$$

With $\boldsymbol{V}_{R,ik}$ and $\boldsymbol{V}_{M,ik}$ specified in terms of parameters like $\sigma^2_{R,k}$ and $\sigma^2_{M,k}$ (and possibly others describing correlation in $\boldsymbol{V}_{R,ik}$), (14) provides a complete description of the interindividual variability impacting how concentrations of "type k" on a specific individual i arise.

The foregoing discussion has focused on a single concentration k. The fact that an entire vector $\boldsymbol{Y}_{ij}$ of $c \geq 1$ concentrations may be measured at each t_{ij} introduces a further complication when $c > 1$. As we noted in Section 4.2, we may conceptualize a representation like that in Figure 4; for *each* k; i.e., for each k there are possible true realized concentration profiles and measurement errors, where the measurement errors for each k arise from different assay procedures. A standard assumption is that the measurement errors committed by different assay procedures when quantifying different samples (e.g., from exhaled air and venous blood)

are completely unrelated. Hence, if we consider the vector $\boldsymbol{e}_{M,ij} = (e_{M,ij1}.e_{M,ij2}, \ldots, e_{M,ijc})'$ consisting of the "measurement error" deviations at time t_{ij}, its covariance matrix will be diagonal, with diagonal elements $(\sigma^2_{M,1}, \sigma^2_{M,2}, \ldots, \sigma^2_{M,c})$ corresponding to the measurement error variances for each assay. However, as we noted in Section 4.2, it could well be that the deviations $e_{R,ijk}$ associated with actual realized concentration profiles for each k at each time point j may be correlated in the sense that "local fluctuations" due to, for example, variation in breathing patterns, and failure of the model to capture perfectly all physiological processes for one concentration may be related to those for another. If we consider the vector $\boldsymbol{e}_{R,ij} = (e_{R,ij1}.e_{R,ij2}, \ldots, e_{R,ijc})'$ consisting of the "realization deviations" at time t_{ij} for all c concentration measures, $\boldsymbol{e}_{R,ij}$ has a nondiagonal covariance matrix whose diagonal elements are $(\sigma^2_{R,1}, \sigma^2_{R,2}, \ldots, \sigma^2_{R,c})$ but whose off-diagonal elements would have to specified according to the analyst's belief regarding how this correlation arises.

What is ordinarily assumed in population PK and PBPK analyses in the literature? The assumptions made are generally not stated explicitly, but in light of our discussion here may be identified straightforwardly. The correlation described in the previous paragraph is generally assumed to be negligible. Whether this is a realistic assumption would need to be critically examined, which could be undertaken informally by calculating sample correlations among the Y_{ijk} across k at each time point; more sophisticated approaches are possible. It is also standard to regard the $e_{R,ijk}$ as being uncorrelated across times t_{ij}. One justification for this assumption that is often given is that the t_{ij} are sufficiently far apart in time relative to the "locality" of the "fluctuations" in a true realized concentration trajectory that correlations among the $e_{R,ijk}$ and hence the Y_{ijk} over time may be disregarded as negligible. Again, such an assumption can be evaluated; Diggle et al. (2001, Ch. 5) discuss diagnostics for this purpose. With these considerations, the Y_{ijk} conditional on $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ are assumed

statistically independent across k; moreover, the matrix $\boldsymbol{V}_{R,ik}$ in (12) and (14) reduces to a diagonal matrix whose diagonal elements are all equal to $\sigma^2_{R,k}$, in which case the matrix $\boldsymbol{V}_{i,k}$ in (14) reduces to a diagonal matrix with diagonal elements given in (10).

This information may be rearranged in an alternative manner consistent with how things are often presented in accounts of population PBPK analyses (e.g., Bois et al., 1996) and the way the individual model is given in (5), as shown next.

4.4 *Popular Individual-Level Model*

As we noted above and at the end of Section 4.2, it is standard in published population PBPK analyses to assume that correlations among deviations, and hence measured concentrations, both over time and across compartments, are negligible, and so can be disregarded in developing the statistical model. This is exactly that, an *assumption*, and it may or may not be correct, as we discuss shortly. Further assumptions are those given on page 25, namely that that the deviations e_{ijk} in (7) and (9) are normally distributed with overall variances σ^2_k as in (10). Under all of these assumptions, the individual-level model is often expressed as follows in published accounts.

Define the matrix $\boldsymbol{\sigma}$ to be a diagonal matrix with diagonal elements $(\sigma^2_1, \ldots, \sigma^2_c)$, where σ^2_k is defined in (10), and off-diagonal elements all equal to zero. Then the assumptions in the last paragraph on the deviations e_{ijk} may be expressed succinctly as the assumption that the vector $\boldsymbol{e}_{ij}$ of these deviations has a $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\sigma})$ distribution, and this is the case for all times $j = 1, \ldots, n_i$. Incorporating this assumption with (5), then, we arrive at the full probability distribution specification for the individual-level model that takes into account beliefs about intraindividual variability, namely

$$\boldsymbol{Y}_{ij} \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \sim \mathcal{N}\{\, \boldsymbol{f}(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i), \boldsymbol{\sigma} \,\} \;\text{ for all }\; j = 1, \ldots, n_i, \tag{15}$$

independently across j. The notation in (15) reminds us that, because we are focusing on individual i only, we are *conditioning on* the exposure pattern $\boldsymbol{E}_i$, physiological measurements $\boldsymbol{\phi}_i$, and unknown PK parameters, treating these as fixed quantities determining how concentration measurements on i arise. An entirely parallel development may be carried out on the log scale as in (8), yielding

$$\log(\boldsymbol{Y}_{ij}) \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \sim \mathcal{N}[\,\log\{\boldsymbol{f}(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)\}, \boldsymbol{\sigma}\,] \;\text{ for all }\; j = 1, \ldots, n_i. \tag{16}$$

Other individual-level models are possible; e.g., like (15) but with intraindividual variability in concentration k proportional to f_k; see Davidian and Giltinan (1995, Ch. 2; 2003, p. 398).

Recall from (10) that the variances σ_k^2 are the sum of two variances characterizing "realization" variability and "measurement error" variability, respectively. In some published analyses, the σ_k^2 are described as being strictly "measurement error variances." This implies the belief that the magnitude of variability due to assay error relative to that due to intraindividual "fluctuations" in the gray curve in Figure 4 is sufficiently large so as to completely dominate the latter's effects, so that effectively

$$\sigma_k^2 = \sigma_{R,k}^2 + \sigma_{M,k}^2 \;\approx\; \sigma_{M,k}^2 \;\text{ for each }\; k. \tag{17}$$

A specification like (15) and assumptions like (17), although routinely made, are not justified unless the analyst believes they are realistic. If they are not, intraindividual variability has been misrepresented, and, as discussed at the outset of this section, this could compromise reliable inferences on the variability in the population.

Are the usual assumptions given at the outset of this section reasonable? They may be in some circumstances and not in others. Assumptions in *any* statistical model should be defensible on scientific grounds, and the consequences of their violation must be understood.

Assumptions such as negligible intraindividual correlations over time or among the c "true" concentrations are often made to simplify the model so that it is easier to fit. However, this is not adequate justification for such an assumption unless it can be demonstrated that the results of analyses are relatively insensitive to it, especially inferences on population parameters. As discussed in Davidian and Giltinan (2003, sec. 2.2.4), there is some evidence that assuming that correlation among realized true concentrations over time is negligible may be reasonable in some cases. Further study in the context of PBPK analysis is required.

4.5 *Implication of PK/PBPK Model Misspecification*

Our development here tacitly assumes that the PBPK model leading to the expression f_k is a reasonable specification for "inherent trajectories," with actual realized concentration profiles varying about them. Thus, the $e_{R,ijk}$ are assumed to represent minor departures from the inherent trajectories due to the inability of any "smooth" deterministic function to capture all the relevant biology. However, if the PBPK model is not sufficiently rich to capture adequately more predominant features of the true "inherent" behavior; e.g., it involves three compartments when four are really needed, then a rather serious PK model misspecification is involved. Here, the f_k may not be acceptable approximations to the "inherent trajectories" of relevant concentrations, and, accordingly, part of what the $e_{R,ijk}$ are representing is in fact *systematic bias*, which the statistical model will interpret wrongly as part of random intraindividual variability. In this case, the meaning and relevance of the PK parameters in the misspecified model may be compromised.

A natural question is whether one can deduce easily whether such a model misspecification has been committed. In some situations, a plot of concentration-time data may reveal clearly that a PK model is grossly inadequate. In other settings, this may not be entirely

obvious, as the model misspecification may not be too great, but still substantial enough to endanger interpretation. Here, the analyst faces the classical problem of distinguishing "signal" from "noise." The fact that the data deviate in a mild but apparently systematic fashion from the inherent trajectories dictated by the model may be due failure of the PBPK model (misrepresentation of the "signal") *or* failure of the statistical model to characterize accurately intraindividual variability, in particular intraindividual *correlation* (misrepresentation of the "noise"). However, based on the data alone, it is impossible to know which is the true explanation! This conundrum can sometimes be resolved by careful consideration of additional scientific information, but it must be understood that the data alone are not sufficient under these conditions distinguish mechanistic model misspecification from statistical model misspecification.

A more detailed discussion of this may be found at the end of Section 2.2.2 of Davidian and Giltinan (2003). In any event, the possibility of PBPK model misspecification substantiates the importance of identifying as realistic a PBPK model as is possible, as population analyses are predicated on its correctness.

4.6 *Interindividual Variability Modeling Considerations*

We have already reviewed the basic considerations underlying the stage 2 population model (6). Here, we mention some additional issues. In some population PK analyses, the covariance matrix $\mathbf{\Sigma}$ representing interindividual variability and covariability of the $\boldsymbol{\theta}_i$ is taken to be a diagonal matrix, with, as before, diagonal elements $(\Sigma_1^2, \Sigma_2^2, \ldots, \Sigma_p^2)$. This implies a belief that there are *no associations* in the population among PK parameters. This is almost always a highly unrealistic assumption often made to simplify fitting of (5)-(6). Restricting $\mathbf{\Sigma}$ to be diagonal when it is not misrepresents fairly dramatically the nature of variability of

$\boldsymbol{\theta}_i$ in the population, forcing any covariability that truly exists to be accommodated by the model in some other way. A result is that the estimates of the population variances Σ_ℓ^2 as well as aspects of intraindividual variability can be flawed.

As noted at the end of Section 3, a refined population analysis attempts deduce whether some of the population variability in the $\boldsymbol{\theta}_i$ can be attributed to systematic associations between the $\boldsymbol{\theta}_i$ and individual attributes. For example, Jonsson et al. (2001) discuss a study of methyl chloride, which is metabolized by the enzyme glutathione S-tranferase T1, for which a genetic polymorphism has been shown; on this basis, individuals in the population may be classified as "conjugators" or "nonconjugators," and conjugator status is clearly systematically associated with metabolism. Similarly, Mezzetti et al. (2003) recorded information on gender and race/ethnicity, and age group, factors which might be thought to explain some population variability in kinetics.

To discuss the required generalization of the population model (6), first note that an equivalent way to represent (6) is

$$\log(\boldsymbol{\theta}_i) = \boldsymbol{\mu} + \boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \quad i = 1, \ldots, N, \tag{18}$$

where, as before $\boldsymbol{\theta}_i$ and hence $\boldsymbol{\mu}$ and $\boldsymbol{b}_i$ have length p. In (18), the $\boldsymbol{b}_i$ may be viewed as a mean zero "deviation" random vector that represents where the PK parameters for individual i "sit" in the population relative to the population mean $\boldsymbol{\mu}$ and are referred to as *random effects*. (18) is reminiscent of a "regression model," where $\log(\boldsymbol{\theta}_i)$ plays the role of the "dependent variable" and the "mean" is the same for all individuals.

If individual attributes $\boldsymbol{A}_i$ have been collected, it is natural to think that the mean of the "dependent variable" $\log(\boldsymbol{\theta}_i)$ might be systematically associated with values of $\boldsymbol{A}_i$, as in a conventional regression model. As a simple example, suppose that the single attribute

"conjugator status" has been collected, represented by $A_i = 0$ for non-conjugators and $A_i = 1$ for conjugators. We may wish to allow the possibility that the mean value of $\boldsymbol{\theta}_i$ in the subpopulation of non-conjugators, $\boldsymbol{\mu}_0$, say, is different from that in the subpopulation of conjugators, $\boldsymbol{\mu}_1$, say, although the variability about the mean in each subpopulation is similar. This may be accommodated by modifying (18) to

$$\log(\boldsymbol{\theta}_i) = \boldsymbol{\mu}_0 A_i + \boldsymbol{\mu}_1(1 - A_i) + \boldsymbol{b}_i. \tag{19}$$

If we define $\boldsymbol{\beta}$ to be the vector of length $2p$ stacking $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ together, then we may interpret interpret $\boldsymbol{\beta}$ as a "regression parameter" in the "regression model" (19). This idea can be extended to models incorporating multiple attributes simultaneously. Writing "regression models" in terms of "regression parameters" and "random effects" for this purpose is standard in the population PK literature (e.g., Davidian and Giltinan, 2003, sec. 2.2.1). A key objective of population PK analyses for pharmaceutical agents is to "build" such regression models to identify and include attributes systematically associated with population variability (Maitre et al., 1991; Mandema et al., 1992; Davidian and Gallant, 1992). We do not consider more general population models like (19) in this article. We do emphasize, however, that $\boldsymbol{\mu}$ in (18) and indeed $\boldsymbol{\beta}$ in (19) are population parameters and thereby are of central interest in population analysis. In Section 5.1, we make additional comments on the interpretation of $\boldsymbol{\mu}$, which extend to parameters like $\boldsymbol{\beta}$ in fancier models.

The population model (6), and, equivalently, (18), make a specific assumption about how $\log(\boldsymbol{\theta}_i)$ are distributed in the population, namely, that they take on their values in the population according to a normal distribution. This, and indeed the assumption of any specific probability distribution for this purpose, imposes a structure with certain features for the population that may not always be consistent with the true structure. For example,

the normal distribution is *unimodal* and *symmetric*, having a single peak representing values that are most likely to be seen, with more extreme values less but equally likely due to symmetry. Some populations may have a structure that is *bimodal*, with two peaks representing possible subpopulations attributable to a feature that is unknown or unmeasured, or may be *asymmetric*. Inference on population parameters like the population mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ could be compromised if the probability distribution chosen for the population model is unrealistic. In the context of drugs, models that replace the normality assumption by a less restrictive one including many different possible probability distributions and methods for fitting them have been proposed (e.g., Mallet, 1986; Schumitzky, 1991; Davidian and Gallant, 1992; Rosner and Müller, 1994; Müller and Rosner, 1997).

4.7 *Summary*

Integrating the considerations in Sections 4.2– 4.4 with the basic hierarchical model (5)-(6), a fully specified model involves (i) a probability distribution describing how, conditional on exposure $\boldsymbol{E}_i$, physiological measurements $\boldsymbol{\phi}_i$, and PK parameters $\boldsymbol{\theta}_i$ for each individual $i = 1, \ldots, N$, measured concentrations arise on each i, which embodies an explicit set of assumptions about intraindividual variability; and (ii) a probability distribution describing how PK parameters $\boldsymbol{\theta}_i$ occur and vary in the population, thereby explicitly representing interindividual variability. A key objective is to estimate the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ that describe (ii), as well as other quantities. In Section 5, we discuss how these probability distributions play a role in the development of estimation methods for this purpose.

Intuitively, concentration measurements on the same individual share common features simply because they are from the same individual, whose individual PK parameters dictate how they arise. More formally, concentration measures on the same individual are *correlated*

at the *level of the population* in that concentrations on the same individual may tend to be "high" or "low" together relative to those on other individuals because of this shared dependence. (This is a different phenomenon from the *intraindividual correlation* discussed in Sections 4.2–4.4, which is relevant even if only a single individual is of interest.) The hierarchical model automatically implies that this is the case, because all Y_{ijk} for individual i depend on the common $\boldsymbol{\theta}_i$; see Davidian and Giltinan (2003, sec. 2.2.4) for more. Valid methods for population analysis based on the hierarchical statistical model thought to give rise to the data naturally take this feature into account.

4.8 *Variability vs. Uncertainty*

It is important to recognize that specification of the hierarchical statistical model is made independently of any estimation method for implementing the model; e.g., using the Bayesian inferential approach becoming increasingly population in this regard, discussed in Section 5.8. The term "nonlinear mixed effects model," for example, refers specifically to models like this that describe how longitudinal data in a study of multiple individuals are thought to arise and may be used independently of reference to a particular estimation method. There has been some confusion about this in the traditional population PK literature, where this term has been construed to refer also to the class of estimation techniques based on approximation to a relevant likelihood function, reviewed in Section 5.5. As a result, other estimation techniques like the "two-stage" methods reviewed in Section 5.4 have been interpreted mistakenly as being irrelevant to "nonlinear mixed effects models" when they are really just another way to fit them, as we demonstrate shortly.

It should be clear from this section that the term *variability* as it is usually used in the literature on population analysis refers to the fact that biological entities and observations

we might make on them vary naturally. In some cases, we may be able to control the extent of variability; for example, if the magnitude of variability due to measurement error is substantial, it may be possible to develop a more precise assay technique that can offer a reduction in intraindividual variability attributable to measurement error. Other sources of variability are dictated by nature, and thus we are "stuck with them." If we are interested in a particular population of humans, for example, interindividual variability in PK in this population is whatever it is. A goal of population analysis is to characterize the nature and magnitude of this "uncontrollable" variability.

Understanding variability has another important implication, having to do with the concept of *uncertainty*. Classically, given we have a statistical model that we believe gives an approximately realistic description of the true data generating process, uncertainty has to do with how well we can learn about, i.e., estimate, quantities of interest in the model based on data in the face of variability. In population analysis, the main quantities of interest are population parameters, and thus a population analysis must also quantify the uncertainty with which these parameters are estimated. Quantifying uncertainty is important because it gives us a sense of the faith we should be attaching to the results of an analysis. An estimate of a population parameter for which the associated uncertainty is large is of little value. Thus, formally assessing and quantifying uncertainty is a requirement for any analysis of data leading to estimates based on a statistical model.

Intuitively, if variability in the processes giving rise to data is large, we might expect to obtain estimates that have high associated uncertainty. This premise is formalized in the classical way of quantifying uncertainty in frequentist statistical thinking through the notion of *sampling variability*, which we review briefly. In this way of thinking, the sample of data we actually obtain after conducting an experiment is viewed as only one of many possible such

samples (with the same sample sizes N and n_i) that we could have ended up obtaining. Each potential such sample would lead to an estimate for the population parameter of interest, where it is taken as given that there is a fixed, true value of this parameter that characterizes the population; e.g., the average value of $V_{\max}$ in the population is 0.2 mg/min, say. If all of these possible estimates *vary* considerably across all these potential samples, then, informally, we would feel "uncertain" that the estimate we obtained from our sample is a reliable reflection of the true value of the population parameter, because if we had ended up with a different sample, the estimate could have been quite different. On the other hand, if all possible estimates do not vary too much, then we would feel fairly "certain" that our estimate is reliable, because another sample would have given something similar. This suggests that we can quantify uncertainty by quantifying the variability across the estimates that could be obtained from all possible samples, referred to as *sampling variability*. The *sampling distribution*, the probability distribution that describes how the possible estimates take on their values across all potential samples, gives a formal picture of this sampling variability. An estimate of the standard deviation of the sampling distribution, often called a *standard error*, is the measure used when a single number is desired to quantify the extent of such sampling variability; the entire sampling distribution gives a complete picture.

For valid estimation methods for fitting complex statistical models like (5)-(6), the exact form of the sampling distribution associated with the estimates of population and other parameters is generally impossible to determine analytically. However, it may be approximated. One way to approximate the sampling distribution is by a mathematical approximation that is typically reasonable for "large" sample sizes ("asymptotic" approximation); here, the approximate sampling distribution is usually a normal distribution whose mean is equal to the true value of the parameter being estimated. Another way to approximate sampling

distributions is by the simulation procedure known as the *bootstrap*.

It is well-established that the sampling distribution and the extent of sampling variability it involves depend on two things: the extent of variability in the processes giving rise to data and the sample size. In the hierarchical model (5)-(6), then, phenomena contributing to sampling variability are the magnitudes of intra- and interindividual variability leading to the data, the number of individuals N sampled from the population, and the number of observation times n_i on each. As discussed above, these sources of variability, especially interindividual variability, are whatever they are. Thus, this dictates that if we wish to achieve smaller sampling variability and thus reduce uncertainty for inferences on population parameters, our primary option is to conduct larger studies. In most population studies of humans, interindividual variability tends to be the predominant source of variability; thus, it should come as no surprise that studies with larger N will lead to sampling distributions of estimates of population parameters that have smaller sampling variability.

The foregoing discussion reviews the differences between variability and uncertainty and the connection between these concepts from the classical point of view. The term uncertainty is also sometimes used to refer to other aspects of population analysis. For example, if there are components of or assumptions in our statistical model that we feel are shaky, e.g., three compartments vs. four compartments for the PBPK model or negligible vs. nonnegligible correlation, then this also makes us uncertain, but, as we have discussed, the consequences are more likely to manifest as estimates that may be misleading because they do not estimate what we think they do, and this would be the case even if there were only modest variability. In this article, we use the term uncertainty to refer specifically to the quality of estimates in an approximately "correct" model, as described above, and use "model misspecification" to refer to these other possible pitfalls.

It is a common misconception that hierarchical statistical models are by nature "Bayesian" and that Bayesian methods are required to characterize the variability and uncertainty inherent in analyses based on the model. As we discuss in the next section, quantifying variability and characterizing uncertainty is possible in any of the estimation methods for the hierarchical model (5)-(6). The Bayesian approach takes a different view of representing uncertainty, which we review in Section 5.8 and contrast with the classical perspective above. The Bayesian approach has some particularly attractive features for population analysis using PBPK models, as we highlight, which has made it a focus of great recent interest.

5 Parameter Estimation

In this section, we review a number of methods that have been used for fitting the model (5)-(6) to data like those described in Section 4.1. Our discussion focuses primarily on estimation of the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, although it turns out that estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ must be carried out jointly with estimation of other quantities, such as the intraindividual variances σ_k^2, $k = 1, \ldots, c$, in (10). For convenience, we refer to all of the quantities to be estimated collectively as $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})$, adopting the notation $\boldsymbol{\sigma}$ at the end of Section 4.4 to summarize the intraindividual variances. In discussing estimation methods, we follow the conventional use of the term "parameter" in statistics to refer to an unknown quantity to be estimated; thus, both the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as well as $\boldsymbol{\sigma}$ are "parameters" to be estimated the statistical model, although our main interest is in the population parameters. Some of these methods have traditionally been used in population analyses based on empirical PK models for drugs, for which the numbers of PK parameters are considerably smaller than in PBPK models and for which the numbers of individuals N in studies may be considerably larger than in human exposure studies. As we discuss later,

many of these have some limitations for analyses based on more complex PBPK models.

Readers wishing a "big picture" understanding of the methods free of most technical details need read only selected parts of this section. All readers should read Sections 5.1, 5.2, 5.7, 5.8, and 5.10. Readers desiring a more technical description of the methods will wish to review Sections 5.3–5.6 and 5.9; these sections can be skipped by "big picture" readers. More formal accounts of these methods are in Davidian and Giltinan (1995; 2003, sec. 3). Although estimation of the population parameters is usually the key objective, methods to "estimate" individual PK parameters $\boldsymbol{\theta}_i$ associated with specific individuals is also possible within the hierarchical model framework. Section 5.11, which is also rather technical, reviews methods for doing so, and can be skipped by readers not interested in the details.

5.1 *What Not To Do, And Why*

Before we discuss valid methods, we mention a few key points.

An important function of a statistical model is to provide a sound basis for developing estimation methods; as we will see shortly, methods for population analysis follow directly from consideration of the hierarchical model. However, some analysts have been tempted to "simplify" population analysis by estimating $\boldsymbol{\mu}$ via an ad hoc approach that does not respect the structure of the model and data. The fact that the observations Y_{ijk} arise from different individuals i is effectively ignored, and the "nonlinear regression model"

$$Y_{ijk} = f_k\{t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \exp(\boldsymbol{\mu})\} + \epsilon_{ijk} \tag{20}$$

is fitted by ordinary least squares (OLS), treating $\boldsymbol{E}_i, \boldsymbol{\phi}_i$ as fixed quantities and $\boldsymbol{\mu}$ as the "regression parameter." In (20), note that $\log(\boldsymbol{\theta}_i)$ is replaced by $\boldsymbol{\mu}$, as if the (log) PK parameters for all individuals were all equal to their mean values in the population. Moreover,

the entire structure of intra- and interindividual variability is ignored, as the OLS method regards all of the ϵ_{ijk} in (20) as statistically independent with mean zero and the same variance; thus, this approach ignores the correlation among concentration measurements on the same individual discussed in Section 4.7. A way to estimate $\boldsymbol{\Sigma}$ is not even evident. In light of the considerations in Section 4, this approach clearly has no defensible rationale. For these reasons, Sheiner and Beal (1980) refer to this as the "naive pooled data" method and present evidence demonstrating its poor performance in the context of empirical PK models.

In some toxicokinetic studies, the data are available only in *aggregated* form. One way this arises is when individuals are observed longitudinally at the same time points; however, the individual concentration measurements Y_{ijk} are not available. Rather, only sample averages and sample variances across individuals are reported at each time point (e.g., Covington et al., 2006; Hack et al., 2006l Marino et al., 2006). An analysis purporting to estimate $\boldsymbol{\mu}$ (and $\boldsymbol{\Sigma}$) might be based on treating these as "data." For example, assuming all individuals received the same exposure $\boldsymbol{E}$ and writing the sample averages over i as $\overline{Y}_{jk}$ at times t_j, say, one possibility would be to "fit" a "model" like

$$\overline{Y}_{jk} = f_k\{t_j, \boldsymbol{E}, \overline{\boldsymbol{\phi}}, \exp(\boldsymbol{\mu})\} + \epsilon_{jk},$$

where $\overline{\boldsymbol{\phi}}$ is average of the known physiological measurements across i. This cannot lead to a credible estimate of $\boldsymbol{\mu}$, as follows. The way in which we believe concentration measurements to arise does not change: each individual still has his/her own PK parameters $\boldsymbol{\theta}_i$ and physiology $\boldsymbol{\phi}_i$ governing how his/her concentrations arise. Thus, the hierarchical model is still a valid representation of how the individual concentration measurements that go into the aggregated sample averages and variances arise. Thus, any valid method for estimating the population parameters should respect the underlying hierarchical model. Trying to use

the sample averages in this ad hoc way does not satisfy this requirement.

Another form of aggregated data comes about in animal studies involving serial sacrifice. Here, each individual animal contributes only a single concentration measurement, so that sample averages and variances at different time points do not involve repeated measurements from the same individuals; i.e., the data are cross-sectional. Nonetheless, as before, the hierarchical model still provides a description of how even a single concentration measurement arises for any animal (governed by that animal's $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_i$) and hence ad hoc estimation methods that do not respect its structure are suspect.

These considerations indicate that, ideally, in either of these aggregated data situations, a valid estimation method for the population parameters must use the hierarchical model that describes how the data that were aggregated arose as a starting point to derive a corresponding model to describe the aggregated data. This implied aggregated data model would then be an appropriate basis for estimation of population parameters. Derivation of such a model is possible in principle, and we discuss this further in Section 6.1. Some analysts wonder whether it is really necessary to do this as opposed to carrying out an ad hoc analysis as above; how "bad" can the latter really be? A comprehensive study of this issue in the context of PBPK population has not to our knowledge been conducted. It may well be that in certain situations reasonable ad hoc estimates of the population parameters may be obtained; however, as such circumstances have not been characterized, we discourage this practice. It is also worth noting that, regardless of the quality of the population parameter estimates so obtained, it is not clear how to characterize faithfully the uncertainty associated with them without reference to an appropriate statistical model (a point discussed further by Sheiner and Beal, 1980, in the context of the "naive pooled data" method above).

To further understand why the ad hoc methods in this section are invalid, it is useful

to appreciate the interpretation of the population parameters, especially $\boldsymbol{\mu}$. From (6) and Section 4.6, we have already emphasized that the correct interpretation of $\boldsymbol{\mu}$ is as the average, or "typical" value of the unknown (log) PK parameters in the population. If we substitute $\boldsymbol{\mu}$ for the unknown log PK parameters in the PK model, as in (20), the resulting expression is *not* the same thing as the "typical" or average concentration profile. That is, $\boldsymbol{\mu}$ does *not* have the interpretation as the value of the log PK parameter leading to "average concentrations." To appreciate this distinction in a simple case, consider Figure 3 and pretend that the 12 individuals in the plot are the entire population and that the depicted profiles are the "inherent trajectories" for each individual, given by the one compartment model (2),

$$\frac{k_{a,i}D}{V_i(k_{a,i}-Cl_i/V_i)}\{\exp(-Cl_it/V_i)-\exp(-k_{a,i}t)\},\quad B\equiv 1, \tag{21}$$

where $\boldsymbol{\theta}_i=(k_{a,i},Cl_i,V_i)$ are the PK parameters for individual i. The "typical" or average concentration profile is found by averaging (21) across all individuals. On the other hand, substituting $\boldsymbol{\mu}=(\mu_{k_a},\mu_{Cl},\mu_V)'$, say, for $\log(\boldsymbol{\theta}_i)$ in (21), gives

$$\frac{e^{\mu_{k_a}}D}{e^{\mu_V}\left(e^{\mu_{k_a}}-e^{\mu_{Cl}}/e^{\mu_V}\right)}\{\exp(-e^{\mu_{Cl}}t/e^{\mu_V})-\exp(-e^{\mu_{k_a}}t)\}, \tag{22}$$

which clearly is not the same. In the first case, the nonlinear functions (21) of (log) $\boldsymbol{\theta}_i$ are averaged directly, whereas in (22) the "averaging" is over the $\log(\boldsymbol{\theta}_i)$ prior to insertion of these into the PK model. Thus, (22) has the interpretation instead as the "inherent concentration profile corresponding to an individual who has the average values of the (log) PK parameters in the population," which is different from the "average of individual concentration profiles across the population." In fact, there may be no real individual in the population with all of his/her (log) PK parameters exactly equal to the "typical" or average value $\boldsymbol{\mu}$! In view of this, the practice of inserting the estimate of $\boldsymbol{\mu}$ into the PK model and plotting the resulting

concentration profiles, as in (22), is a rather meaningless display from the point of view of population analysis. See, Davidian and Giltinan (2003, sec. 2.4) for more discussion of this admittedly subtle but important point.

Finally, because the dimension p of the vector of unknown PK parameters $\boldsymbol{\theta}_i$ can be large in PBPK models, and there is scant information in the data on some of the components, a common practice in some analyses involving PBPK models is to hold these components fixed to literature values that are the same for all individuals and estimate only those parameters "of interest" (a recent example is Yokley et al., 2006). For example, if the focus is on understanding metabolism, literature values may be substituted for all elements of $\boldsymbol{\theta}_i$ except $V_{\max,i}$ and $K_{m,i}$ in (3), with the population analysis then focusing on estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ pertaining to these parameters only. There are conceptual problems associated with this practice, to which we return in the sequel.

5.2 *Classical Methods for Estimation of Population Parameters*

In the previous section, we noted that valid methods for population analysis must be predicated on the hierarchical statistical model (5)-(6) and the analyst's assumptions about its components. A natural basis for classical frequentist statistical inference that respects the form of a statistical model thought to give rise to the data is the method of *maximum likelihood*, which provides a basis for estimating parameters in the model. We now describe the basic idea in the context of the hierarchical model (5)-(6).

To implement maximum likelihood estimation for population analysis based on the hierarchical model, one must write down the so-called *likelihood function* for all parameters to be estimated, which in this case are $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})$. Loosely speaking, the likelihood function describes formally the probability of "seeing the data we saw," which is viewed as depending

on the values of the parameters. For the hierarchical model, the likelihood function is thus found by writing down an expression for this probability based on the probability distributions assumed at both the individual-level and population model stages; i.e., (6) and (15) or (16). In Section 5.3, the form of the likelihood function $L(\boldsymbol{\Omega})$, say, so derived is given explicitly in (25) and is quite complex, as discussed further shortly. The idea underlying maximum likelihood estimation is to take as the "best estimates" for the parameters in the model, $\boldsymbol{\Omega}$ here, as the values that maximize the likelihood function $L(\boldsymbol{\Omega})$. That is, the *maximum likelihood estimates* are taken as the values corresponding to the highest probability of having ended up with the actual data observed. Thus, to estimate $\boldsymbol{\Omega}$ and thereby the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we maximize $L(\boldsymbol{\Omega})$ in $\boldsymbol{\Omega}$.

Standard theory for maximum likelihood provides an approximation to the sampling distribution associated with the resulting estimate $\widehat{\boldsymbol{\Omega}}$ that is reasonable if N is "sufficiently large" (and, of course, if the statistical model is approximately correct and not misspecified). The sampling distribution is an approximate multivariate normal distribution with mean equal to the true value of $\boldsymbol{\Omega}$ and *sampling covariance matrix* that characterizes the sampling variability. This distribution may be used to quantify the (classical) uncertainty associated with the resulting estimate of $\boldsymbol{\Omega}$ (and thus those of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$). The square roots of the diagonal elements of the covariance matrix provide approximate estimated standard errors. The off-diagonal elements of the sampling covariance matrix quantify the extent to which determination of the estimate for one parameter depends on another, and is a critical part of uncertainty assessment; see Bernillon and Bois (2000) for discussion.

As noted above, the form of $L(\boldsymbol{\Omega})$ turns out to be complicated, and its evaluation involves the need to carry out N intractable, high-dimensional integrations. This renders its maximization a very difficult numerical problem, as discussed in more detail in Section 5.3.

As a consequence, methods for population analysis have been developed that attempt to *approximate* the results of maximizing $L(\boldsymbol{\Omega})$, thus yielding estimated values $\widehat{\boldsymbol{\Omega}}$ that are only approximations to the value that would be obtained if maximization of $L(\boldsymbol{\Omega})$ itself were computationally feasible, with corresponding approximations to the sampling distributions. The so-called *two-stage* methods are based on estimating $\boldsymbol{\theta}_i$ for each individual i separately based on the individual-level model (5) using standard methods such as least squares or maximum likelihood for individual data. These estimates are then treated as "data" for estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in particular. Another class of methods is the *first order* methods, various versions of which are based on approximating $L(\boldsymbol{\Omega})$ by something more tractable. Both classes of methods have seen heavy use in population analyses for pharmaceutical agents; technical descriptions and descriptions of the precise manner in which they approximate maximization of $L(\boldsymbol{\Omega})$ are given in Sections 5.4 and 5.5. Approaches to maximizing $L(\boldsymbol{\Omega})$ directly are discussed in Section 5.6. Readers uninterested in the details should proceed to Section 5.7.

5.3 *Maximum Likelihood Estimation**

Here, we present the form of the likelihood function $L(\boldsymbol{\Omega})$ for the hierarchical model, which demonstrates clearly the complexity involved in its practical maximization.

Writing as before $\boldsymbol{Y}_i$ to denote the random vector of all concentration measurements on individual i, the likelihood for $\boldsymbol{\Omega}$ given the observed data, $L(\boldsymbol{\Omega})$, may be based on the joint probability density function of all the $\boldsymbol{Y}_i$, $i = 1, \ldots, N$, given the collection of $(\boldsymbol{E}_i, \boldsymbol{\phi}_i)$, $i = 1, \ldots, N$, which, by the assumed statistical independence across individuals, factors into the product of the N probability density functions for the individual $\boldsymbol{Y}_i$ given $(\boldsymbol{E}_i, \boldsymbol{\phi}_i)$. So we must derive these densities based on the assumptions of the statistical model.

Focusing now on individual i, from the frequentist perspective, $\boldsymbol{\theta}_i$ is an unknown random

vector. Thus, it is "integrated out" from the likelihood function. To discuss estimation methods in a manner consistent with their presentation in the traditional population PK and statistical literature, we invoke an equivalent way of writing the hierarchical model (5)-(6) in which the PBPK model is reexpressed ("reparameterized") in terms of unknown

$$\boldsymbol{\gamma}_i = \log(\boldsymbol{\theta}_i), \quad \boldsymbol{\gamma}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \text{as} \quad \boldsymbol{f}(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i). \tag{23}$$

Adopting this reexpression of the model, letting $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i;\, \boldsymbol{\sigma})$ be either the normal or lognormal probability density function corresponding to the assumptions (15) or (16), and letting $\varphi(\boldsymbol{\gamma}_i;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the p-variate normal probability density function with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the probability density function for $\boldsymbol{Y}_i$ given $(\boldsymbol{E}_i, \boldsymbol{\phi}_i)$ is

$$p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i;\, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \int p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i;\, \boldsymbol{\sigma})\, \varphi(\boldsymbol{\gamma}_i;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{\gamma}_i. \tag{24}$$

The likelihood function is then given as the product of (24) over i, namely

$$L(\boldsymbol{\Omega}) = \prod_{i=1}^{N} p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i;\, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{N} \int p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i;\, \boldsymbol{\sigma})\, \varphi(\boldsymbol{\gamma}_i;\, \boldsymbol{\mu}, \boldsymbol{\Sigma})\, d\boldsymbol{\gamma}_i, \tag{25}$$

where in (25) $\boldsymbol{y}_i$, $i = 1, \ldots, N$, correspond to the actual observed values of $\boldsymbol{Y}_i$, $i = 1, \ldots, N$, once the study is conducted. To estimate $\boldsymbol{\Omega}$ and thereby the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, we maximize $L(\boldsymbol{\Omega})$ in $\boldsymbol{\Omega}$. As noted in Section 5.2, standard theory for maximum likelihood provides an approximation to the sampling distribution associated with the resulting estimate $\widehat{\boldsymbol{\Omega}}$ that is reliable if N is "sufficiently large."

The maximization in principle may be carried out using standard optimization software. A major stumbling block is that the N p-dimensional integrals in (25) are analytically intractable, as the integrands are very complicated, nonlinear functions of $\boldsymbol{\gamma}_i$, and these integrals are high dimensional. Although in principle they can be approximated using numerical

integration techniques (quadrature or Monte Carlo), this can be computationally burdensome for p even as small as 3, as each integral must be numerically evaluated repeatedly at each iteration of the optimization algorithm used to find the maximizer of (25). For PBPK population analyses, the dimension of p makes this approach daunting. Moreover, an additional complication is that the system of differential equations must be solved numerically for each individual at each time point within each optimization iteration. It is thus not surprising that traditional methods for population analysis seek instead to approximate in other ways or otherwise circumvent these integrals, detailed in the next two sections.

5.4 *Two-Stage Methods**

A common misconception is that this approach is somehow unrelated to the hierarchical model, but it may indeed be viewed as an attempt to approximate the integrals in (25). When sufficient data on each individual are available to fit the individual-level model (5) to each individual's data separately; i.e., estimate $\boldsymbol{\gamma}_i$ ($\boldsymbol{\theta}_i$) for individual i based on his/her observed data $\boldsymbol{y}_i$ using a least squares or maximum likelihood method for individual data, one may obtain individual-specific estimates $\widehat{\boldsymbol{\gamma}}_i$ for each $i = 1, \ldots, N$. "Two-stage" methods use these as "data" for estimating $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

A naive approach is to pretend that the $\widehat{\boldsymbol{\gamma}}_i$ are exact "stand-ins" for the unknown $\boldsymbol{\gamma}_i$ and estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ by the sample average and sample covariance matrix of the $\widehat{\boldsymbol{\gamma}}_i$; i.e.,

$$\widehat{\boldsymbol{\mu}} = N^{-1} \sum_{i=1}^{N} \widehat{\boldsymbol{\gamma}}_i, \quad \widehat{\boldsymbol{\Sigma}} = (N-1)^{-1} \sum_{i=1}^{N} (\widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\mu}})(\widehat{\boldsymbol{\gamma}}_i - \widehat{\boldsymbol{\mu}})'.$$

The problem with this approach, which has been called the "Standard Two-Stage" method in the population PK literature (Steimer et al., 1984), is that the $\widehat{\boldsymbol{\gamma}}_i$ are not equal to the $\boldsymbol{\gamma}_i$; instead, they are uncertain estimates of the $\boldsymbol{\gamma}_i$ in the sense discussed in Section 4.8.

The result is that although $\widehat{\boldsymbol{\mu}}$ can give a reasonable estimate of $\boldsymbol{\mu}$, the $\widehat{\boldsymbol{\Sigma}}$ cannot give an accurate reflection of the true interindividual population variability because the $\widehat{\boldsymbol{\gamma}}_i$ also involve additional sampling variability due to the fact that they estimate, but are not exactly equal to, $\boldsymbol{\gamma}_i$ (see Davidian and Giltinan, 1995, sec. 5.3.1).

The naive approach fails because it does not refer back correctly to the structure of the data as represented by the hierarchical model. The sampling variability in each $\widehat{\boldsymbol{\gamma}}_i$ comes about because of the intraindividual variability at the individual level, which is characterized by the individual-level model (5). If one refers to this model, under its assumptions on intraindividual variability, standard sampling theory for least squares or maximum likelihood methods for obtaining $\widehat{\boldsymbol{\gamma}}_i$ yields that conditional on $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$, if the n_i are sufficiently large,

$$\widehat{\boldsymbol{\gamma}}_i \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \,\dot{\sim}\, \mathcal{N}(\boldsymbol{\gamma}_i, \boldsymbol{C}_i), \quad \text{for each} \;\; i = 1, \ldots, N, \tag{26}$$

where "$\dot{\sim}$" means "approximately distributed as." (26) says that, treating $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ as fixed, $\widehat{\boldsymbol{\gamma}}_i$ has a p-variate normal sampling distribution whose sampling variability is approximated by the covariance matrix $\boldsymbol{C}_i$, which in general depends on $\boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i$ as well as $\boldsymbol{\sigma}$. One may substitute $\widehat{\boldsymbol{\gamma}}_i$ and an estimate for $\boldsymbol{\sigma}$ in $\boldsymbol{C}_i$ to obtain $\widehat{\boldsymbol{C}}_i$ and the further approximation

$$\widehat{\boldsymbol{\gamma}}_i \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i \,\dot{\sim}\, \mathcal{N}(\boldsymbol{\gamma}_i, \widehat{\boldsymbol{C}}_i), \quad i = 1, \ldots, N. \tag{27}$$

(27) thus takes into account, albeit approximately, the effects of intraindividual variability as represented in (5). To estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, (5) is then combined with the population model (6), expressed in the form (23). It may be shown that, together, (27) and (23) imply

$$\widehat{\boldsymbol{\gamma}}_i \,|\, \boldsymbol{E}_i, \boldsymbol{\phi}_i \,\dot{\sim}\, \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \widehat{\boldsymbol{C}}_i), \quad i = 1, \ldots, N; \tag{28}$$

note that $\boldsymbol{\theta}_i$ has been "integrated out," so that (28) conditions on just $\boldsymbol{E}_i, \boldsymbol{\phi}_i$ as for the full likelihood analysis based on (25). Estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ proceeds by fitting the approximate

"statistical model" (28) to the "data" $(\widehat{\boldsymbol{\gamma}}_i, \widehat{\boldsymbol{C}}_i)$, $i = 1, \ldots, N$, by maximum likelihood. Practical implementation is discussed in Davidian and Giltinan (1995, Ch. 5; 2003, sec. 5.4.2); Steimer et al. (1984) referred to one way to do this as the "Global Two-Stage" method. An approximate normal sampling distribution associated with the resulting estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ may be obtained in the usual way, pretending that (28) is the "true" statistical model.

This reasoning suggests informally that this approach may be viewed as an attempt to approximate the integrals in the full-blown likelihood function (25). Intuitively, the use of the $\widehat{\boldsymbol{\gamma}}_i$ as "data" distills the actual data on i down to a summary based on the individual-level model, which is then combined with the population model; see Davidian Giltinan (2003, sec. 3.2) for more discussion. Thus, two-stage estimation is a valid way to fit the hierarchical model; in fact, it is approximately equivalent when N and the n_i are "large" to the widely used "first order conditional" method (e.g., Demidenko, 2004, sec. 8.10), discussed next.

5.5 *First Order and Related Methods**

Another class of techniques is based on approximating the likelihood function $L(\boldsymbol{\Omega})$ in (25) by approximating $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, $i = 1, \ldots, N$, directly. Methods in this class are by far the most popular and widely used for traditional population PK analysis. There are several ways to justify the two main types of methods; we demonstrate one popular way and refer the reader to the references at this end of this section for details and other derivations.

Consider for definiteness the stage 1 individual-level model (5). The usual derivation takes $c = 1$, so that $\boldsymbol{Y}_{ij}$ in (5) is one-dimensional; i.e., only one concentration is measured at each t_{ij}, but it is easily extended to $c > 1$, as follows. With the PBPK model expressed

in terms of $\boldsymbol{\gamma}_i = \log(\boldsymbol{\theta}_i)$, we can stack up (5) for $j = 1, \ldots, n_i$ for all cn_i observations as

$$\boldsymbol{Y}_i = \begin{pmatrix} \boldsymbol{Y}_{i1} \\ \boldsymbol{Y}_{i2} \\ \vdots \\ \boldsymbol{Y}_{in_i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}(t_{i1}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i) + \boldsymbol{e}_{i1} \\ \boldsymbol{f}(t_{i2}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i) + \boldsymbol{e}_{i2} \\ \vdots \\ \boldsymbol{f}(t_{in_i}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\gamma}_i) + \boldsymbol{e}_{in_i} \end{pmatrix} = \boldsymbol{f}_i(\boldsymbol{\gamma}_i) + \boldsymbol{e}_i, \tag{29}$$

where $\boldsymbol{e}_i$ is the random vector of all cn_i intraindividual deviations and $\boldsymbol{f}_i(\boldsymbol{\gamma}_i)$ is the vector of length cn_i with the PBPK models stacked up and dependence on the t_{ij}, $\boldsymbol{E}_i$, and $\boldsymbol{\phi}_i$ suppressed for brevity. Following assumptions on the deviations e_{ijk} made according to the considerations in Section 4.3, $\boldsymbol{e}_i$ will have a covariance matrix $\boldsymbol{U}_i(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})$, say, summarizing assumptions made on intraindividual variances for and correlations among all the e_{ijk}. If intraindividual variances are proportional to f_k, as discussed after (16), this could depend on the PK parameters, so we allow dependence of $\boldsymbol{U}_i$ on $\boldsymbol{\gamma}_i$ to be consistent with the literature.

With representation (29) of the entire intraindividual model, we may now demonstrate the argument usually attributed to Beal and Sheiner (1982) leading to the so-called "first-order method" for estimation of $\boldsymbol{\Omega}$. The argument depends on a first-order Taylor series of the model, and is rather technical; the result of the argument is presented in (32) below.

Let $\boldsymbol{U}_i^{1/2}(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})$ be the "square root" matrix of $\boldsymbol{U}_i(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})$ such that $\boldsymbol{U}_i^{1/2}(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})\boldsymbol{U}_i^{1/2}(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})' = \boldsymbol{U}_i(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})$. If $\boldsymbol{\epsilon}_i$ is a random vector that has a $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ distribution, where $\boldsymbol{I}$ is the identity matrix with ones on the diagonal and zeroes everywhere else, then $\boldsymbol{e}_i$ can be written equivalently as $\boldsymbol{e}_i = \boldsymbol{U}_i^{1/2}(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})\boldsymbol{\epsilon}_i$ and (29) can be written as

$$\boldsymbol{Y}_i = \boldsymbol{f}_i(\boldsymbol{\gamma}_i) + \boldsymbol{U}_i^{1/2}(\boldsymbol{\gamma}_i, \boldsymbol{\sigma})\boldsymbol{\epsilon}_i = \boldsymbol{f}_i(\boldsymbol{\mu} + \boldsymbol{b}_i) + \boldsymbol{U}_i^{1/2}(\boldsymbol{\mu} + \boldsymbol{b}_i, \boldsymbol{\sigma})\boldsymbol{\epsilon}_i. \tag{30}$$

Following the literature, in the second part of (30), we have replaced $\boldsymbol{\gamma}_i = \log(\boldsymbol{\theta}_i)$ with its equivalent representation $\boldsymbol{\gamma}_i = \boldsymbol{\mu} + \boldsymbol{b}_i$ as in (18), with $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. Taking a linear Taylor

series of this expression about $\boldsymbol{b}_i = \mathbf{0}$, its mean, in both terms that involve $\boldsymbol{b}_i$ and ignoring the term depending on the "cross-product" $\boldsymbol{b}_i\boldsymbol{\epsilon}_i$ yields

$$\boldsymbol{Y}_i \approx \boldsymbol{f}_i(\boldsymbol{\mu}) + \boldsymbol{Z}_i(\boldsymbol{\mu})\boldsymbol{b}_i + \boldsymbol{U}_i^{1/2}(\boldsymbol{\mu}, \boldsymbol{\sigma})\boldsymbol{\epsilon}_i, \tag{31}$$

where $\boldsymbol{Z}_i$ is the matrix whose elements are partial derivatives of the components of $\boldsymbol{f}_i(\boldsymbol{\gamma}_i)$ with respect to $\boldsymbol{\gamma}_i$. Note that both $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$ are normally distributed random vectors with mean $\mathbf{0}$ that are involved in (31) in an simple additive, linear way, which leads to the resulting approximation for $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i;\, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The upshot of the argument is that $\boldsymbol{Y}_i$ conditional on $\boldsymbol{E}_i, \boldsymbol{\phi}_i$ has approximately a cn_i-variate normal probability distribution with mean and covariance matrix

$$\boldsymbol{f}_i(\boldsymbol{\mu}) \quad \text{and} \quad \boldsymbol{Z}_i(\boldsymbol{\mu})\boldsymbol{\Sigma}\boldsymbol{Z}_i(\boldsymbol{\mu}) + \boldsymbol{U}_i(\boldsymbol{\mu}, \boldsymbol{\sigma}). \tag{32}$$

This implies that we may approximate $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i;\, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ by the normal probability density with mean and covariance matrix (32), thereby circumventing the integrals in (25). The "first order" method of estimating $\boldsymbol{\Omega}$ substitutes this approximation in the first expression for $L(\boldsymbol{\Omega})$ in (25), and $\boldsymbol{\Omega}$ is then estimated by maximizing this "approximate likelihood." This method gained widespread popularity in the 1980s among pharmacokineticists through its implementation in the software package `nonmem` (`http://www.globomaxservice.com/nonmem.htm`), which includes a numerical differential equation solver. Other software, such as the SAS procedure `nlmixed`, also implements this method. A related fitting method is available in the SAS macro `nlinmix` available at `http://support.sas.com`. Galecki (1998) developed a SAS program called `nlmem` that merges a differential equation solver with the `nlinmix` macro to fit models for which f_k are not analytically tractable.

Although this method historically generated considerable excitement, inspection of (32) reveals a drawback. The probability distribution $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i;\, \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ describes how the

data arise "on average" across individuals, because the individual PK parameters $\boldsymbol{\gamma}_i$ have been "integrated out." Thus, the mean vector for $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ has the interpretation as "the average of individual concentration profiles across the population." As we cautioned against in Section 5.1, this method approximates this average by the "inherent concentration profile corresponding to an individual who has the average value $\boldsymbol{\mu}$ of the log PK parameters in the population." As we discussed, these two quantities are not at all the same, raising legitimate concern that the resulting estimate for $\boldsymbol{\mu}$ will be flawed.

A second class of approaches is based on a better approximation to "the average of individual concentration profiles across the population," which is given in (33) below. The poor approximation of the first order method arises because the Taylor series (30) is taken in $\boldsymbol{b}_i$ about its mean $\mathbf{0}$, which is rather crude. A better approximation is obtained by taking the Taylor series about something "closer" to $\boldsymbol{b}_i$. A natural contender is the so-called *empirical Bayes* "estimate" of $\boldsymbol{b}_i$, $\widehat{\boldsymbol{b}}_i$, say, discussed in Section 5.11, which can be derived straightforwardly from the hierarchical model and depends on the data and value of $\boldsymbol{\Omega}$. Assuming availability of $\widehat{\boldsymbol{b}}_i$, a linear Taylor series in (30) about $\boldsymbol{b}_i = \widehat{\boldsymbol{b}}_i$ leads to approximating $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ by a cn_i-variate normal density with mean and covariance matrix

$$\boldsymbol{f}_i(\boldsymbol{\mu} + \widehat{\boldsymbol{b}}_i) - \boldsymbol{Z}_i(\boldsymbol{\mu} + \widehat{\boldsymbol{b}}_i)\widehat{\boldsymbol{b}}_i \quad \text{and} \quad \boldsymbol{Z}_i(\boldsymbol{\mu} + \widehat{\boldsymbol{b}}_i)\boldsymbol{\Sigma}\boldsymbol{Z}_i(\boldsymbol{\mu} + \widehat{\boldsymbol{b}}_i) + \boldsymbol{U}_i(\boldsymbol{\mu} + \widehat{\boldsymbol{b}}_i, \boldsymbol{\sigma}). \tag{33}$$

To maximize (25) with this normal density substituted for $p_i(\boldsymbol{y}_i|\boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\sigma}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, popular implementations iterate between (i) updating $\widehat{\boldsymbol{b}}_i$ for each $i = 1, \ldots, N$ with $\boldsymbol{\Omega}$ held fixed at a current estimate and (ii) obtaining the next current estimate of $\boldsymbol{\Omega}$ by maximizing (25) in $\boldsymbol{\Omega}$ holding the $\widehat{\boldsymbol{b}}_i$, $i = 1, \ldots, N$, fixed at their values from (i). This approach goes by various names; we call it the "first-order conditional" method consistent with how it is referred to in the `nonmem` software. There are variations on this theme in various software packages,

including `nonmem`, the SAS procedure `nlmixed`, the SAS macros `nlinmix` and `nlmem`, and the R suite of functions `nlme` (Pinheiro and Bates, 2000); a version of the latter with a built-in differential equation solver, `nlmeode`, is also available.

For either the first order or first order conditional methods, the sampling distribution associated with the estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ is obtained by acting as if the approximate likelihood is the "true" likelihood. This gives an approximate normal sampling distribution, from which standard errors may be obtained in the usual way. Empirical studies have shown that estimates of uncertainty based on this approximate sampling distribution are very reliable in general as long as N is not too small.

There is an extensive literature on methods based on such approximations; see Beal and Sheiner (1982), Lindstrom and Bates (1990), Vonesh and Carter (1992), Wolfinger (1993), Davidian and Giltinan (1995, Ch. 6; 2003, sec. 3.3), Pinheiro and Bates (1995); Vonesh (1996), Vonesh and Chinchilli (1997), Wolfinger and Lin (1997), among many others. There are also many other software packages implementing these methods, many dedicated to PK analysis, such as `winnonmix` (`http://www.pharsight.com/products/prod_winnonmix_home.php`).

5.6 *"Direct" Maximization of the Likelihood**

The methods in the previous two sections are based on analytical manipulations to approximate the likelihood (25) so as to avoid the integrations involved. As mentioned at the outset, it is of course possible to use numerical approximations to the integrals instead, which is often referred to as a "direct" or "exact" method by practitioners, although it still involves a (numerical) approximation. There are software packages that implement variations on this theme, including the SAS procedure `nlmixed`, `usc*pack` (Jellife, et al., 1996), and `monolix` (`http://www.math.u-psud.fr/ lavielle/monolix/logiciels`). As noted earlier, this can

be computationally demanding when the dimension p of $\boldsymbol{\gamma}_i$ is large.

5.7 *Remarks on Practical Implementation*

An important consideration for population analysis using any of the methods discussed so far is that it is *hard*. The likelihood function $L(\boldsymbol{\Omega})$ (25) or even approximations to it for a typical PK population model is a complicated, highly nonlinear function of the parameters $\boldsymbol{\Omega}$. Thus, it can be challenging to maximize, and good starting values for the optimization routines used are essential. This is especially true when the dimension p of the unknown $\boldsymbol{\theta}_i$ is large, as is typical for PBPK models, because, even if the integrals are circumvented analytically, the number of components in $\boldsymbol{\Omega}$ that must be estimated is also large. A further problem in this case that measurements on only a few accessible concentrations may not contain sufficient information to achieve *structural identification*; i.e., to *identify* the behavior of some of the numerous PK parameters in the PBPK model, particularly those in deeper compartments. Accordingly, the data may not have the requisite information to identify the population mean and variance associated with some of these PK parameters. This makes it tempting to set the corresponding elements of $\boldsymbol{\theta}_i$ equal to the same, known value derived from the literature for all individuals, as discussed at the end of Section 5.1 in order reduce the dimension of what has to be estimated so as to "make it work."

It should be obvious at this point that this last practice, which is often followed in the analysis of individual data using PBPK models, is dangerous if our goal is to learn realistically about interindividual variability in PK in the population. From a biological perspective, the idea that some PK parameters vary in the population while others are exactly the same for everyone is ludicrous indeed, but this is precisely what this practice assumes. Intuitively, assessment of the variability across, and even the population means values of, the remaining

PK parameters that are not fixed in this way will be flawed if the variability of others that really do vary in the population is artificially set to zero, because the contributions to the overall variability in the data by the latter are not properly acknowledged. As remarked earlier, adopting simplifying assumptions solely to facilitate fitting of the model is not good analysis practice and can lead to misleading results.

Of course, if there is little information in data on some of the PK parameters and hence the components of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ corresponding to them, it is not satisfying to abandon any attempt at analysis, either. A compromise that is sometimes used in traditional population analysis is to set the components of $\boldsymbol{\mu}$ corresponding to such parameters to a fixed literature value, but to still allow variability in the parameter across the population. The Bayesian approach, discussed next, offers a possibly better way of handling this difficulty in some circumstances. However, it is important to recognize that Bayesian inference it is not a completely "magic bullet," either, as we will remark shortly.

5.8 *Bayesian Inference*

So far, the methods we have discussed are based, through approximations of various kinds, on classical maximum likelihood, with uncertainty quantified by approximate (normal) sampling distributions derived from large sample theory. The Bayesian approach to statistical inference is based on an alternative point of view to the repeated-sampling-based, frequentist perspective on uncertainty, although under certain conditions the two approaches can be reconciled. There are several features of the Bayesian framework that make it an attractive basis for population analysis with PBPK models, which explains the upsurge of interest in this approach. Key among these is that the Bayesian paradigm provides an explicit mechanism for introducing outside information, e.g., based on previous studies carried out by the

analyst or reported in the literature, or even "educated guesses," into an analysis, which has special appeal. However, there are also potential pitfalls that must be appreciated.

Technically, the Bayesian approach to inference does not distinguish between "random vectors" that represent data that might be collected or unknown, unobservable quantities like $\boldsymbol{\theta}_i$ (or $\boldsymbol{\gamma}_i$) and "parameters" such as the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ whose values are to be estimated; all model components are considered "random" in that they are taken to have probability distributions. Of course, in a population modeling and analysis context, as alluded to throughout and noted explicitly in Section 4.8, we certainly believe that there are *fixed*, *true* values of quantities like $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ describing the population in which we are interested, so it might seem confusing to think that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ would have probability distributions of possible values! This is reconciled by the view that the Bayesian approach takes on probability: "probability" is viewed as a *measure of uncertainty*. Thus, for our purposes, probability distributions associated with population parameters, for example, may be interpreted as expressing the degrees of uncertainty about their (fixed) values. A comprehensive reference on Bayesian inference explicating this view is Gelman et al. (2004).

How does this work? The ingredients of a Bayesian analysis follow from *Bayes rule*. We first demonstrate in a generic situation where we have a statistical model depending on a random variable Y, say, representing "data" that might be collected and a parameter μ, say, perhaps representing the mean of all possible values Y could take on. If we view μ as a "random variable," too, we may envision the joint probability density of Y and μ that describes the probabilities with which they take on their values together, which we write as $p(y, \mu)$ and that by standard results factors as

$$p(y, \mu) = p(y|\mu)p(\mu). \tag{34}$$

In (34), $p(y|\mu)$ is the (conditional) probability density of Y treating μ as fixed, which thus describes how the data would take on their values when the mean of their possible values is fixed at μ. That is, $p(y|\mu)$ represents our belief in how the data take on their values, as we ordinarily specify in a statistical model. The probability density $p(\mu)$ is called the *prior density*. The prior may be thought of as specifying our belief, or uncertainty, about the value of μ *in the absence of information from the data* Y. If $p(\mu)$ is very spread out, this indicates that, before we have seen the data, we are pretty uncertain about what the value of μ might be. If $p(\mu)$ is concentrated around a particular value, this reflects that we are fairly certain what the value of μ is before we have even collected data. Bayes rule provides a mechanism for "updating" our assessment of this uncertainty *once we have seen data*. Define $p(\mu|y)$ to be the probability density that reexpresses our uncertainty *given* that we have now seen actual data whose value turned out to be y. Bayes rule states that

$$p(\mu|y) = \frac{p(y|\mu)p(\mu)}{p(y)}, \quad p(y) = \int p(y|\mu)\, p(\mu) d\mu). \tag{35}$$

In (35), $p(\mu|y)$ is called the *posterior density*, with the above interpretation. The posterior consequently is the fundamental object by which uncertainty about the parameter μ is summarized following the observation of data. The goal of a data analysis is hence to obtain the posterior for quantities of interest like μ, where a prior for μ has been specified.

Note that (35) involves the need to evaluate an integral. In complex statistical models like the hierarchical model (5)-(6), the forms of the posteriors for parameters like $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are complicated, involving numerous high-dimensional integrals; readers wishing to see a technical demonstration should read Section 5.9. The need to carry out complex integrations like these in useful but complex statistical models was a major roadblock to the widespread application of Bayesian methods until the advent of modern computing power and advances

in Markov chain Monte Carlo (MCMC) techniques made it possible to "do" such integration by simulation numerically. MCMC methods are a body of numerical simulation techniques that have many uses, not limited to the implementation of Bayesian inference, although it is for this purpose that they have enjoyed great recent popularity. It is important to recognize that MCMC methods are simply a computational device that proves convenient for calculating posteriors and thus operationalizing Bayesian inference. As such, they are not themselves a "statistical method." Readers desiring an overview of how MCMC methods work in the context of population analysis should read Section 5.9.

5.9 *Population Analysis in a Bayesian Framework**

We demonstrate how Bayesian inference is operationalized for population analysis under the hierarchical model (5)-(6) (or equivalently (18)). From the Bayesian perspective, all of the potential data vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, N$; the unobserved log PK parameters $\boldsymbol{\gamma}_i = \log(\boldsymbol{\theta}_i)$, $i = 1, \ldots, N$; the population parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$; and the additional intraindividual variability parameters $\boldsymbol{\sigma}$ are viewed as random vectors with probability distributions. As in the classical analyses discussed previously, we view the collection of observed $\boldsymbol{E}_i, \boldsymbol{\phi}_i$, $i = 1, \ldots, N$, as fixed and known for the purposes of analysis (so condition on their observed values throughout). Our goal is to obtain suitable posterior densities for the quantities $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}, \boldsymbol{\gamma}_i$, $i = 1, \ldots, N$, given we have observed actual data $\boldsymbol{y}_i$ along with the fixed $\boldsymbol{E}_i, \boldsymbol{\phi}_i$, $i = 1, \ldots, N$.

To implement the Bayesian machinery to this end, we must specify a prior distribution for the parameters $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})$, which is often called a "hyperprior" in the context of population analysis; we denote the prior as $p(\boldsymbol{\Omega}) = p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})$ here. There is no "prior" *per se* attached to the $\boldsymbol{\gamma}_i$, $i = 1, \ldots, N$; as specified in the hierarchical model, the $\boldsymbol{\gamma}_i$ are taken to follow the population model (18). In some Bayesian analyses, the probability distribution

dictated by the population model is sometimes also called a "prior," but we refrain from this terminology in order to maintain the view of the hierarchical model that this distribution dictates how individuals in the population arise regardless of the type of analysis (Bayesian or classical) undertaken, so is part of the "data generating" model. Writing for short $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_i,\, i = 1, \ldots, N\}$, $\boldsymbol{y} = \{\boldsymbol{y}_i,\, i = 1, \ldots, N\}$, and $(\boldsymbol{E}, \boldsymbol{\phi}) = \{(\boldsymbol{E}_i, \boldsymbol{\phi}_i),\, i = 1, \ldots, N\}$, it may be shown that the joint posterior density of all of $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}, \boldsymbol{\gamma}_i,\, i = 1, \ldots, N$ is given by

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}, \boldsymbol{\gamma} | \boldsymbol{y}, \boldsymbol{E}, \boldsymbol{\phi}) = \frac{\displaystyle\prod_{i=1}^{N} p(\boldsymbol{y}_i | \boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\gamma}_i; \boldsymbol{\sigma})\, \varphi(\boldsymbol{\gamma}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})}{\displaystyle\int\int\int\int \prod_{i=1}^{N} p(\boldsymbol{y}_i | \boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\gamma}_i; \boldsymbol{\sigma})\, \varphi(\boldsymbol{\gamma}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})\, d\boldsymbol{\gamma} d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}\, d\boldsymbol{\sigma}}. \tag{36}$$

Comparing to (35), note that the probability densities corresponding to the intra- and interindividual components of the hierarchical model appear in the position corresponding to our belief on how the data arise, as expected.

The posterior density (36) is the *joint* posterior density of all of $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}, \boldsymbol{\gamma}_i,\, i = 1, \ldots, N$. Conventionally, assessment of uncertainty for particular parameters, such as $\boldsymbol{\mu}$ is made by examining the posterior density for that parameter alone. Operationally, to obtain the individual posterior density of $\boldsymbol{\mu}$ given the data $\boldsymbol{y}$ and $(\boldsymbol{E}, \boldsymbol{\phi})$, $p(\boldsymbol{\mu} | \boldsymbol{y}, \boldsymbol{E}, \boldsymbol{\phi})$, say, we must integrate (36) with respect to $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\gamma}$; i.e., we need to calculate

$$p(\boldsymbol{\mu} | \boldsymbol{y}, \boldsymbol{E}, \boldsymbol{\phi}) = \frac{\displaystyle\int\int\int \prod_{i=1}^{N} p(\boldsymbol{y}_i | \boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\gamma}_i; \boldsymbol{\sigma})\, \varphi(\boldsymbol{\gamma}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})\, d\boldsymbol{\gamma}\, d\boldsymbol{\Sigma}\, d\boldsymbol{\sigma}}{\displaystyle\int\int\int\int \prod_{i=1}^{N} p(\boldsymbol{y}_i | \boldsymbol{E}_i, \boldsymbol{\phi}_i; \boldsymbol{\gamma}_i; \boldsymbol{\sigma})\, \varphi(\boldsymbol{\gamma}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma})\, p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})\, d\boldsymbol{\gamma} d\boldsymbol{\mu}\, d\boldsymbol{\Sigma}\, d\boldsymbol{\sigma}}. \tag{37}$$

Clearly this is a monumental task that cannot be done analytically! Nonetheless, in principle if we could obtain the form of the posterior density $p(\boldsymbol{\mu} | \boldsymbol{y}, \boldsymbol{E}, \boldsymbol{\phi})$, we would have a complete picture of uncertainty about the value of the population parameter $\boldsymbol{\mu}$ (and similarly for

$\boldsymbol{\Sigma}$ and $\boldsymbol{\sigma}$). The modal value, mean, or median value of the posterior is usually reported as the "estimate" of the parameter. The standard deviation of the posterior is reported when a single numerical measure of uncertainty is desired. It is important to recognize that uncertainty here is in the sense described above and hence has a different interpretation from the sampling-based measures used classically (but see below).

As noted in Section 5.8, these complex integrations may be implemented numerically by using MCMC techniques. A complete description of how MCMC methods work in the context of implementing Bayesian population PK analysis is beyond our scope here, and we refer the reader to accounts by Wakefield et al. (1994), Davidian and Giltinan (1995, Ch. 8), Bennett et al. (1996), Bernillon and Bois (2000), and Gelman et al. (2004, Ch. 11) for some of the nuts and bolts issues and to Rosner and Müller (1994), Bois et al. (1996), Gelman et al. (1996), Wakefield (1996), Müller and Rosner (1997), Jonsson and Johanson (2001ab), Jonsson et al. (2001), Jonsson and Johanson (2002), and Mezzetti et al. (2003) for accounts of population analyses. A brief description is as follows.

The end product of implementation of Bayesian population analysis by using MCMC methods to perform the integrations are large samples of values of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\gamma}_i$, $i = 1, \ldots, N$, simulated from their joint posterior. Using these samples, one may construct numerically any feature of the posterior one desires. For example, the sample of $\boldsymbol{\mu}$ values represents a sample of possible values arising from the posterior density $p(\boldsymbol{\mu}|\, \boldsymbol{y}, \boldsymbol{E}, \boldsymbol{\phi})$ in (37), and a visual estimate of this can be obtained by, e.g., plotting a histogram for the sample. The Bayesian "estimate" of $\boldsymbol{\mu}$ can be found as the mean, median, or mode of the the values, and a numerical summary of the uncertainty in it can be obtained as their standard deviation. Because samples from the joint posterior (36 are in fact obtained, information on the joint uncertainty (e.g., the extent to which uncertainty about $\boldsymbol{\mu}$ is associated with that about $\boldsymbol{\Sigma}$)

is available.

These samples are obtained by an iterative scheme that produces a simulated sequence of values of all of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\gamma}_i$, $i = 1, \ldots, N$, such that the sequence forms a Markov chain that, technically speaking, has stationary distribution that is the same as the joint posterior of interest ((36) here). After a sufficient number of iterations to ensure that this property is realized, the values in the sequences may be viewed as samples from the posterior. Details of how to decide what constitutes a "sufficient number," how to construct the values in the sequence using techniques like the Metropolis and Metropolis-Hasting algorithms are reviewed in the references cited above.

Software to carry out these calculations is available. An add-on to the popular package `bugs` (Bayesian inference Using Gibbs Sampling) called `pkbugs` (PKBugs, 2004) has been developed for fitting population PK models with a focus on those relevant to pharmaceutical studies. The package `MCSim` (Bois et al., 2003) is tailored to fitting population PBPK models, and includes an differential equation solver. `MCSim` has been used in a number of published population analyses with success (e.g., Jonsson and Johanson; 2001ab; Jonsson et al., 2001; Jonsson and Johanson, 2002).

5.10 *Contrasting Bayesian and Classical Methods*

We now are in a position to comment on some of the advantages of the Bayesian approach as well as some of the potential pitfalls. A comprehensive discussion is clearly not possible here, so we touch only on a few key points.

The prior distribution represents a natural mechanism for incorporating information on the values of population and other parameters available in the literature or in previously collected data. The prior can also be used to impose constraints that force the values of

parameters like $\boldsymbol{\mu}$ to stay in biologically plausible ranges through choices of priors that do not include values outside these ranges as possibilities (thus representing perfect certainty that they do not fall outside these ranges). When little information is available, prior distributions that are "flat" are often used. In fact, if the prior contains no information (so we are "infinitely uncertain" *a priori* about plausible values for the parameters), the posterior will be proportional to the data generating model that defines the likelihood, Bayesian and likelihood analyses will be approximately equivalent, and the classical sampling distribution associated with a likelihood-based estimate and the posterior will coincide (i.e., will be approximate normal distributions; see Gelman et al., 2004, Ch. 4). When there is prior information available, the shape and extent of spread in the prior may be chosen to reflect what is known from the literature or previous experiments about parameters in what are called "informative" priors. The shape of the resulting posterior will be dictated by the merging of the prior and data. Commonly, priors are specified separately for each parameter, partly because it may come from different sources and partly for simplicity of implementation. Formally, this treats the prior information on each parameter as statistically independent, and means that the prior for population analysis is specified as the product $p(\boldsymbol{\Omega}) = p(\boldsymbol{\mu})P(\boldsymbol{\Sigma})p(\boldsymbol{\sigma})$.

A routine criticism of Bayesian analysis in general is in regard to the choice of the prior. When the prior is chosen for "convenience" because it simplifies computations or makes some calculations analytically tractable, such criticism can be justified. To deflect this criticism, it is essential that priors be constructed faithfully using the best scientific information available. Why is there concern about choice of prior? Because the posterior that is the central tool for inference depends on the prior, it is fair to be concerned that the results (inferences based on the posterior) may be sensitive to how the prior was chosen. Two different priors can lead

to different posteriors and hence to possibly different conclusions. When the goal is to make policy recommendations, this is less than desirable. Thus, any Bayesian analysis should include an investigation of such sensitivity, as discussed in the aforementioned references.

Just as with the classical approaches to population analysis, implementation of the Bayesian approach can be computationally challenging and intensive, and there is no guarantee that the sequence of simulated values will exhibit good behavior; e.g., the sequence should "stabilize" or converge to produce values that are consistent with the true posterior density. In some circumstances, such convergence can be very slow or elusive and needs to be monitored carefully using diagnostic procedures that have been developed for this purpose. Some analysts will run separate sequences and inspect the results for comparability of the resulting simulated posteriors. Thus, recalling Section 5.7, computational hurdles arise in population analysis regardless of the statistical inferential approach, and the analyst must be prepared to inspect carefully the results rather than adopting them readily.

A source of tension between adherents to the classical frequentist and Bayesian approaches is in regard to the role of data in inferences. In the frequentist approach, the process by which the data arise, including the sample sizes chosen when conducting an experiment, are central to drawing conclusions and defining and quantifying uncertainty, and uncertainty will be large and the entire exercise infeasible when there are little data or little information in them. In the Bayesian approach, the data are only part of the picture, and this has led some to argue that large sample sizes are somehow not necessary for Bayesian analysis. However, with little data (or data that carry little information), a posterior can still be obtained, and its form in this case will be dictated largely by the form of the prior, which opens the possibility for inferences and uncertainty evaluations that may be almost entirely predicated on the prior if it is constructed to be "informative." This reinforces the

concern about choice of the prior, particularly when it is specified based on summary measures in the literature from narrowly defined populations or by extrapolation from animal experiments. Some frequentists argue that when historical data are used to construct priors, an alternative to a Bayesian analysis is to include these historical data with those from the current experiment and to develop an overall statistical model for how *all* of them arise. They also note that it is possible to impose constraints on parameter values, in an albeit different way, in classical likelihood-based analyses.

There is no easy resolution to this debate! Most pragmatic analysts agree that, as long as its scope and limitations are appreciated, the Bayesian approach is natural in the context of population analysis. In practice, the debate may be somewhat moot; when the data contain "good" information, the results of classical and Bayesian analyses are often closely aligned; compare Davidian and Gallant (1992) and Wakefield (1996), which report classical and Bayesian population analyses of a PK study of the anti-arrhythmic drug quinidine.

5.11 *Estimation of Individual PK Parameters**

As noted previously, the main focus of population analysis is on the population, not on specific individuals. However, there may be situations where inference on PK parameters for individuals in a study is desired. "Estimates" of the PK parameters $\boldsymbol{\gamma}_i$ or equivalently $\boldsymbol{\theta}_i$ for specific individuals i may be obtained from any of the methods we have discussed.

From the Bayesian approach, these arise naturally. A posterior for each of the $\boldsymbol{\gamma}_i$, $i = 1, \ldots, N$, is the probability density

$$p(\boldsymbol{\gamma}_i | \boldsymbol{y}, \boldsymbol{E}, \boldsymbol{\phi}), \quad i = 1, \ldots, N. \tag{38}$$

Samples from the posteriors (38) are a byproduct of the analysis implemented by MCMC

simulation. Thus, means (or medians or modes) of these may be used as "estimates" of the $\boldsymbol{\gamma}_i$, and their histogram and other summaries (e.g. standard deviations) quantify uncertainty.

In the classical approach, because the $\boldsymbol{\gamma}_i$ are regarded as random vectors, so on a footing different from that of the fixed population parameters, use of the term "estimate" is considered misplaced, and often the term "predictor" is used instead. It may be argued from this perspective that in fact a "reasonable" such predictor is a so-called *empirical Bayes estimate*; a simple version of the argument is given in Davidian and Giltinan (1995, Ch. 3), and a general discussion of the idea of empirical Bayes inference is discussed by Carlin and Louis (2000). The posterior (38), similar to (37), is the probability density found by integrating the joint posterior with respect to all of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\sigma}$, and the other $\boldsymbol{\gamma}_k$, $i \neq i$, so cannot be found analytically. However, if one treats $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}$ as known constants, one can write down a posterior density of $\boldsymbol{\gamma}_i$ for individual i of the form

$$p(|\boldsymbol{\gamma}_i|\boldsymbol{y}_i, \boldsymbol{E}_i, \boldsymbol{\phi}_i;\ \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}), \tag{39}$$

which depends on the parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma}$ (not "integrated out"). The empirical Bayes approach uses a density like (39) as the basis for inference on random vectors like $\boldsymbol{\gamma}_i$ in the hierarchical models. Because (39) depends on $\boldsymbol{\Omega} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\sigma})$, an estimate of $\boldsymbol{\Omega}$ is substituted in (39). The mode of (39); i.e., the value of $\boldsymbol{\gamma}_i$ maximizing (39), is then given as the predictor or "estimate." The term "empirical Bayes" is thus used to emphasize that, while this has a Bayesian flavor, "empirical" estimates of $\boldsymbol{\Omega}$ are substituted rather than being "integrated out" as in the fully Bayesian approach.

Because $\boldsymbol{b}_i = \boldsymbol{\gamma}_i - \boldsymbol{\mu}$ from (18), equivalently, one may find an empirical Bayes estimate for $\boldsymbol{b}_i$, which is used in the "first order conditional" method, as discussed in Section 5.5.

6 Implications and Extensions

6.1 *Extensions*

Our presentation has been in terms of a specific version of the hierarchical statistical model that forms the basis for population PK analysis assuming that the data are in a certain form where c different concentrations are measured on each individual i at the same time points t_{ij}, $j = 1, \ldots, n_i$. This situation was chosen to provide a concrete basis for describing the critical features of intraindividual and interindividual that must be considered in specifying the statistical model, but it does not mean that the model cannot be extended and generalized to other settings. We have already remarked (Section 5.1) that the basic hierarchical structure

At the end of Section 4.6, we mentioned one such extension, that of including individual attributes in the population model to take into possible account and uncover their systematic associations with PK parameters in the population. It is also possible to allow the time points at which each of the c concentrations are measured on each individual to be different, as, for example, it may be logistically difficult to collect both blood and breath samples according to the same schedule. The same considerations on intraindividual correlation discussed in Sections 4.2–4.4 would of course apply.

Another important extension of the hierarchical model is to this situation where the same individuals may be observed on more than one "occasion," e.g., where each is exposed to three different concentrations of a hazard agent in three different exposure chamber episodes separated by suitable washout periods, and longitudinal concentration measurements are taken for each. Although the same PBPK model is good approximation to the PK behavior across all these occasions, it is plausible that the PK parameters describing this behavior may vary somewhat within the same individual over time, reflecting natural physiological

variation. In this case, letting ℓ index the occasion, the model may be extended to allow PK parameters $\boldsymbol{\theta}_{i\ell}$, say, for each occasion within individual i. Karlsson and Sheiner (1993) discuss modeling considerations for this situation.

In some instances, only aggregated data may be available, as discussed in Section 5.1, and the objective is to estimate population parameters on their basis. We noted in that section and reiterate here that, even though the data that are available are aggregated, the process by which such aggregated values arise still ultimately is represented by the hierarchical statistical model. In particular, the concentration measurements for each individual i that have been aggregated may be thought to arise for each individual according to this hierarchy, where individual i's concentration measurements are governed by his/her $\boldsymbol{\phi}_i$ and $\boldsymbol{\theta}_i$ values. Accordingly, we emphasize again that valid estimation of population parameters based on aggregated data should be based on methods obtained by assuming that individual concentration measurements arise according to the hierarchical model. As suggested in Section 5.1, this would involve developing a model for aggregated data based on the hierarchical model. For example, suppose that concentration measurements for the kth compartment for $i = 1, \ldots, N$ individuals have been summarized at each of n (common) times t_j, $j = 1, \ldots, n$, by the sample average and standard deviation across N. At the jth time, this means that the available "data" are the sample average and standard deviation of the Y_{ijk}, given by

$$m_{jk} = N^{-1} \sum_{i=1}^{N} Y_{ijk}, \quad s_{jk} = \left\{ (N-1)^{-1} \sum_{i=1}^{N} (Y_{ijk} - m_{jk})^2 \right\}^{1/2}; \tag{40}$$

the Y_{ijk} themselves are not observed. More precisely, then, the task would be to derive based on the hierarchical model for the Y_{ijk} a corresponding model for the probability distributions describing how m_{jk} and s_{jk} would take on their values. Given the complexity of the hierarchical model, this may be analytically and computationally challenging. Well-performing

methods for this problem need to be developed.

Another common challenge in PBPK analysis is that (individual or aggregated) data from different sources may be available, and each data source contains information on different aspects of the kinetics. For example, one data set may involve measured intravenous concentrations that contain information on metabolic parameters, while another may be suitable for estimation of absorption characteristics. It is natural to attempt to estimate population parameters on which each data set has information separately by data set and then combining the estimates post hoc. This practice ignores the possibility that there may be information on other aspects, such as intraindividual variability and other PBPK parameters in some or all available data sets that might be exploited. As for aggregated data, a fruitful approach under these conditions may be to use the hierarchical model as a starting point for describing how data in the various studies arise and then to combine them in an overarching, "meta-analytic" statistical model framework that acknowledges explicitly the different data sources. Wakefield and Rahman (2000) discuss this approach in the context of population analysis for pharmaceutical agents. Further research on such methods for PBPK analysis, which involves additional challenges, is needed.

Any population analysis will only be as reliable as the mechanistic PBPK model and hierarchical statistical model in which is it is embedded. It is natural for an analyst to ask "what did I assume, and what if I am wrong?" As noted in Section 4.5, the overall statistical model may be inappropriate because of misspecification of the PK model, inappropriate assumptions on intra- and interindividual variability, or both. Thus, there is great interest in tools for assessing the suitability of the model components. Several popular techniques for this purpose are available. One widely-used method is to inspect some sort of *information criterion*, which may be thought of roughly as an objective measure of overall model fit

adjusted to take into account the number of quantities estimated. One may fit several competing models that differ in regard to, say, the mechanistic model embedded or complexity of assumptions on intraindividual variability. The model fit yielding the most favorable value of the information criterion might be judged the "best;" however, there are some caveats. We have already noted the difficulty of distinguishing between a poorly chosen PK model and unsuitable statistical assumptions, e.g., on intraindividual correlation. Because information criteria can only judge the "overall fit," it is impossible without further scientific assumptions to identify a "best" mechanistic model or "best" set of intraindividual variability assumptions because of the way these features are "tangled up" together, as described in Section 4.5. Thus, it may be difficult to use such measures to identify an appropriate PBPK model. Given one is willing to assume, for example, that the chosen PBPK model is reasonable, then, fixing this model, one may investigate the statistical modeling assumptions using information criteria as well as diagnostic plots and other measures, some of which are mentioned in (4.3. This often involves some subjectivity interpreting graphical displays, and conclusions can be unduly influenced by features like "unusual" concentration measurements or "outlying" individuals who do not seem to be consistent with the others. In many cases, settling on an an appropriate overall model is a bit of an "art form." Continued research into methods for model selection and diagnosis in the context of population PBPK analysis is welcome.

Our presentation has followed standard accounts in the literature, which treat the physiological measurements $\boldsymbol{\phi}_i$ fixed quantities for the i individuals and carry out the analysis *conditional* on their values for the N sampled in individuals for the purpose of characterizing the distribution of unknown PK parameters ($\boldsymbol{\theta}_i$). Of course, the $\boldsymbol{\phi}$ also are distributed in, and hence vary in, the population; e.g., body weight, lean body mass, and so on clearly

exhibit variability across individuals. If the model is to be used to *simulate* individuals in the population, as described in the next section, it is important that the distribution of the physiological measures also be characterized. In fact, ideally, the entire *joint* distribution of the $\boldsymbol{\phi}$ and $\boldsymbol{\theta}_i$ would be required. See Mentré and Mallet (1994), Rosner and Müller (1994), and Müller and Rosner (1997).

6.2 *Further Analyses*

Our presentation to this point has focused on methods for and issues associated with fitting the hierarchical statistical framework embedding a PBPK model to data, referred to as population analysis. Although the results of a fit, such as estimates of population parameters and estimates of uncertainty for them, are useful and important in their own right, generally they are not the ultimate objective of a population analysis. Indeed, as we noted at the outset, the goal is usually to exploit the fitted model as a basis for further analyses. We now make some brief comments on issues that must be appreciated in this endeavor.

One important goal for risk assessment is to understand the relationship between delivered dose, i.e., concentration of the agent in target tissue, and some outcome or response. As a first step, it is important to understand the nature of the concentrations that might be seen in the population under a particular exposure level/pattern, in particular, the *variability* in delivered doses that might be expected if all individuals were subjected to the same exposure. Intuitively, variability in PK parameters across the population will dictate variability in achieved target tissue concentrations. Thus, a fitted hierarchical model embedding a PBPK model for the agent can form the basis for such an investigation. The population model (6) characterizes PK parameter population variability. Armed with this and a distribution of physiological parameters for the population, one may use Monte Carlo simulation

to draw "virtual" individuals from the population and via the PBPK model to obtain the "inherent" delivered doses dictated by each. A histogram of the resulting simulated delivered doses would provide an empirical view of the population variability in concentrations. Note that Monte Carlo simulation in this context simply refers to making random draws from a probability distribution and is distinct from MCMC techniques for fitting the model.

Note that any subsequent analysis involving estimates of the population parameters will be subject to uncertainty, just as are the estimates of these parameters themselves. Once the parameter estimates are obtained, they cannot be regarded as known fixed quantities; rather it must be appreciated that the uncertainty associated with them will propagate through to the conclusions of any further analyses based on them. Thus, the aforementioned simulation is in fact subject to uncertainty, because the simulated delivered doses are based on the population estimates.

7 Concluding Remarks

In this article, we have tried to give a systematic and detailed account of the formulation of the statistical model framework underlying population analysis and of some the methods that have historically and more recently been used for its implementation. Of necessity, we have not touched on numerous important issues, nor have we been able to provide a comprehensive review of all methods, model extensions, and so on that are possible. We refer the reader to the references for more details. The bibliography for population PK analysis we provide in this article is in no way exhaustive, and we encourage the interested reader to seek out further references, as new advances continue to emerge.

Clearly, there are many open problems in population analysis using PBPK analysis still to be resolved, and we look forward to continued research and advances in this area.

Acknowledgments

This work was supported in part by NIH grants R01-CA085848 and R37-AI031789. The author gratefully acknowledges the excellent suggestions of A. John Bailer, Weihsueh Chiu, and Woodrow Setzer in preparing this article.

Population PK analysis using hierarchical nonlinear models was pioneered by Lewis B. Sheiner (1940-2004) and Stuart L. Beal (1941-2006).

References

Beal, S.L. and Sheiner, L.B. (1982). Estimating population pharmacokinetics. *CRC Critical Reviews in Biomedical Engineering* **8**, 195–222.

Bennett, J.E., Racine-Poon, A., and Wakefield, J.C. (1996). MCMC for nonlinear hierarchical models. In *Markov Chain Monte Carlo in Practice*, ed. W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. New York: Chapman and Hall/CRC Press.

Bernillon, P. and Bois, F.Y. (2000). Statistical issues in toxicokinetic modeling: A Bayesian perspective. *Environmental Health Perspectives* **108**(Suppl), S883–893.

Bois, F.Y., Gelman, A., Jiang, J., Maszle, D., Zeise, L., and Alexeef, G. (1996). Population toxicokinetics of tetrachloroethylene. *Archives of Toxicology* **70**, 347–355.

Bois, F.Y., Jackson, E.T., Pekari, K., and Smith, M.T. (1998). Population toxicokinetics of benzene. *Environmental Health Perspectives Supplements* **104**, 1405–1411.

Bois, F.Y., Maszle, D.R., Revzan, K., Tiller, S., and Yuan, Z. (2003). MCSim: a Monte Carlo Simulation Program, version 5.0.0, available at `http://toxi.ineris.fr/activites/toxicologie_quantitative/mcsim/mcsim.php`

Carlin, B.P. and Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*. Boca Raton: Chapman and Hall/CRC Press.

Covington, T.R., Gentry, P.R., Van Landingham, C.B., Andersen, M.E., Kester, J.E., and Clewell, H.J. (2006). The use of Markov chain Monte Carlo uncertainty analysis to support a Public Health Goal for perchloroethylene. *Regulatory Toxicology and Pharmacology*,
doi:10.1016/j.yrtph.2006.06.008.

Davidian, M. and Gallant, A.R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine. *Journal of Pharmacokinetics and Biopharmaceutics* **20**, 529–556.

Davidian, M. and Giltinan, D. M. (1995).*Nonlinear Models for Repeated Measurement Data*. London: Chapman and Hall.

Davidian, M. and Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, **8** 387–419.

Demidenko, E. (2004). *Mixed Models: Theory and Applications*. New York: Wiley.

Diggle, P.J., Heagerty, P., Liang, K.-Y., and Zeger, S.L. (2001). *Analysis of Longitudinal Data, Second Edition*. Oxford University Press.

Galecki, A.T. (1998). NLMEM: A new SAS/IML macro for hierarchical nonlinear models. *Computer Methods and Programs in Biomedicine* **55**, 207–216.

Gear, C.W. (1971). The automatic integration of ordinary differential equations. *Communications of the Association for Computing Machinery* **14**, 176–179.

Gelman, A., Bois, F., and Jiang, J. (1996). Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* **91**, 1400–1412.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). *Bayesian Data Analysis, Second Edition*. Boca Raton: Chapman and Hall/CRC Press.

Gibaldi, M. and Perrier, D. (1982). *Pharmacokinetics* (2nd edn). New York: Marcel Dekker.

Giltinan, D.M. (2006). Pharmacokinetics and pharmacodynamics. In *Encyclopedia of Biostatistics* (2nd edn), ed. P. Armitage and T. Colton, 000–000 New York: Wiley. DOI: 10.1002/0470011815.b2a07040.

Hack, C.E., Chiu, W.A., Zhao, Q.J., and Clewell, H.J. (2006). Bayesian population analysis of a harmonized physiologically based pharmacokinetic model of trichloroethylene and its metabolites. *Regulartory Toxicology and Pharmacology* **46**. 63–83.

Jellife R.W., Schumitzky, A., van Guilder, M. (eds.) (1996). User Manual for Version 10.6 USC*PACK Collection of PC Programs. Laboratory of Applied Pharmacokinetics, University of Southern California, Los Angeles, California.

Johanson, G., Jonsson, F., and Bois, F. (1999). Development of a new technique for risk assessment using physiologically based toxicokinetic models. *American Journal of Industrial Medicine*, **36**S1, 101–103.

Jonsson, F., Bois, F.Y., and Johanson, G.(2001). Assessing the reliability of PBPK models using data from methylchloride-exposed, non-conjugating human subjects. *Archives of Toxicology* **75**, 189–199.

Jonsson, F. and Johanson, G. (2001a). A Bayesian analysis of the influence of GSTT1 polymorphism on the cancer risk estimate for dichloromethane. *Toxicology and Applied Pharmacology* **174**, 99–112.

Jonsson, F. and Johanson, G. (2001b). Bayesian estimation of variability in adipose tissue blood flow in man by physiologically based pharmacokinetic modeling of inhalation exposure to toluene. *Toxicology* **157**, 177–193.

Jonsson, F. and Johanson, G. (2002). Physiologically based modeling of the inhalation kinetics of styrene in humans using a Bayesian population approach. *Toxicology and Applied Pharmacology* **179**. 35–49.

Jonsson, F. and Johanson, G. (2003). The Bayesian population approach to physiological toxicokinetic-toxicodynamic models – An example using the MCSim software. *Toxicology Letters* **138**, 143–150.

Karlsson, M.O. and Sheiner, L.B. (1993). The importance of modeling inter-occasion variability in population pharmacokinetic analyses. *Journal of Pharmacokinetics and Biopharmaceutics* **21**, 735–750.

Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673–687.

Maitre, P.O., Buhrer, M., Thomson, D., and Stanski, D.R. (1991). A three-step approach combining Bayesian regression and NONMEM population analysis: Application to midazolam. *Journal of Pharmacokinetics and Biopharmaceutics* **19**, 377–384.

Mallet, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika* **73**, 645–656.

Mandema, J.W., Verotta, D., and Sheiner, L.B. (1992). Building population pharmacokinetic/pharmacodynamic models. *Journal of Pharmacokinetics and Biopharmaceutics* **20**, 645–656.

Marino, D.J., Clewell, H.J., Gentry, P.R., Covington, T.R., Hack, C.E., David, R.M., and Morgott, D.A. (2006). Revised assessment of cancer risk to dicholoromethane: Part I Bayesian PBPK and dose-response modeling in mice. *Regulartory Toxicology and Pharmacology* **45**, 44–54.

Mentré, F. and Mallet, A. (1994). Handling covariates in population pharmacokinetics.

International Journal of Bio-Medical Computing **36**, 25–33.

Mezzetti, M., Ibrahim, J.G., Bois, F.Y., Ryan, L.M., Ngo, L., and Smith, T.J. (2003). A Bayesian compartmental model for the evaluation of 1,3-butadiene metabolism. *Applied Statistics* **52**, 291–305.

Müller, P. and Rosner, G.L. (1997). A Bayesian population model with hierarchical mixture priors applied to blood count data. *Journal of the American Statistical Association* **92**, 1279–1292.

Pinheiro, J.C. and Bates, D.M. (1995). Approximation to the log-likelihood function in the nonlinear mixed effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and Splus*. New York: Springer.

PKBugs (2004). *PKBugs: An Efficient Interface for Population PK/PD within WinBUGS*. London: Imperial College School of Medicine.
`http://www.winbugs-development.org.uk/pkbugs/home.html`.

Racine-Poon, A. (1985). A Bayesian approach to nonlinear mixed effects models. *Biometrics* **41**, 1015–1023.

Rosner, G.L. and Müller, P. (1994). Pharmacokinetic/pharmacodynamic analysis of hematologic profiles. *Journal of Pharmacokinetic and Biopharmaceutics* **22**, 499–524.

Rowland, M. and Tozer, T.N. (1995). *Clinical Pharmacokinetics – Concepts and Applications, Third Edition*. Media, PA: Williams and Wilkins.

Schumitzky, A. (1991). Nonparametric EM algorithms for estimating prior distributions. *Applied Mathematics* **45**, 143–157.

Sheiner, L.B. and Beal, S.L. (1980). Evaluation of methods for estimating population phar-

macokinetic paramters. I. Michaelis-Menten model: Routine linical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics* **8**, 553–571.

Sheiner, L.B. and Ludden, T.M. (1992). Population pharmacokinetics/dynamics. *Annual Review of Pharmacology and Toxicology* **32**, 185–209.

Sheiner, L.B., Rosenberg, B., and Melmon, K.L. (1972). Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research* **5**, 441–459.

Sheiner, L.B., Rosenberg, B., and Marathe, V.V. (1977). Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *Journal of Pharmacokinetics and Biopharmaceutics* **5**, 445–479.

Steimer, J.L., Mallet, A., Golmard, J.L., and Boisvieux, J.F. (1984). Alternative approaches to estimation of population pharmacokinetic parameters: Comparison with the nonlinear mixed effect model. *Drug Metabolism Reviews* **15**, 265–292.

U.S. Food and Drug Administration (FDA) (1999). Guidance for Industry: Population Pharmacokinetics. Available at `http://www.fda.gov/CDER/guidance/1852fnl.pdf`.

Vonesh, E.F. and Carter, R.L. (1992). Mixed effects nonlinear regression for unbalanced repeated measures. *Biometrics* **48**, 1–17.

Vonesh, E.F. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika* **83**, 447-452.

Vonesh, E.F. and Chinchilli, V.M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Wakefield, J. (1996). The Bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association* **91**, 271–283.

Wakefield, J.C. and Rahman, N. (2000). The combination of population pharmacokinetic

studies. *Biometrics* **56**, 263–270.

Wakefield, J.C., Smith, A.F.M., Racine-Poon, A., and Gelfand, A.E. (1994). Bayesian analysis of linear and nonlinear population models by using the Gibbs sampler. *Applied Statistics* **43**, 201–221.

Wolfinger, R.D. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791–795.

Wolfinger, R.D. and Lin, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Computational Statistics and Data Analysis* **25**, 465–490.

Yokley, K., Tran, H.T., Pekari, K., Rappaport, S., Riihimaki, V., Rothman, N., Waidyanatha, S., and Schlosser, P.M. (2006). Physiologically-based pharmacokinetic modeling in humans: A Bayesian approach. *Risk Analysis*, doi:10.1111/j.1539-6924.2006.00789.x

Table 1: List of selected symbols. See the text for more detailed descriptions.

Symbol	Meaning
i	Index identifying individiduals sampled from the population
N	Total number of individuals
j	Index identifying time points at which compartment-specific concentrations are measured on the ith individual
n_i	Number of time points at which concentrations are measured on individual i
t_{ij}	The jth time at which concentrations are measured on the ith individual
k	Index identifying the kth compartment on which measured concentrations are available *
c	Number of compartments on which concentrations are measured at each time t_{ij} ($c \leq$ total number of compartments in the model)
$\boldsymbol{E}_i$	Collection of all information on exposure level and duration experienced by individual i
Y_{ijk}	Compartment-specific concentration measurement on compartment k at time t_{ij} on individual i
$\boldsymbol{Y}_{ij}$	Collection of all c compartment-specific concentrations measured at time t_{ij} on individual i
$\boldsymbol{Y}_i$	Collection of all compartment-specific concentrations over all times measured on individual i
$\boldsymbol{\phi}_i$	Collection of physiological measurements on individual i (known, measured quantity)
$\boldsymbol{\theta}_i$	Collection of all unknown PK parameters (possibly transformed and/or rescaled) associated with individual i
$\boldsymbol{A}_i$	Collection of measured, known attributes on individual i, such as gender, ethnicity, genotypic information
$f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$	Expression for the kth compartment-specific (mean) concentration at time t found by solving the PBPK equations under i's exposure $\boldsymbol{E}_i$, physiology $\boldsymbol{\phi}_i$, and PK parameters $\boldsymbol{\theta}_i$

Table 1: List of selected symbols, continued. See the text for more detailed descriptions.

Symbol	Meaning
$\boldsymbol{f}(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$	Collection of expressions for all c compartment-specific (mean) concentrations for individual i at time t
e_{ijk}	Deviation representing how measured concentration Y_{ijk} differs from the expected concentration $f_k(t_{ij}, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$ at time t_{ij} under the PBPK model
$\boldsymbol{e}_{ij}$	Collection of deviations associated with all c compartment-specific concentrations measured at time t_{ij} on individual i
$\boldsymbol{\mu}$	Collection of population means of each element of the unknown PK parameters $\boldsymbol{\theta}_i$ across the pouplation of all individuals (population parameter)
$\boldsymbol{\Sigma}$	Covariance matrix of the elements of the unknown PK parameters across the population of all individuals (population parameter)
Σ^2_ℓ	Variance in the population of individuals of the ℓth PK parameter in $\boldsymbol{\theta}_i$ (and ℓth diagonal element of $\boldsymbol{\Sigma}$)
$e_{R,ijk}$	Part of deviation e_{ijk} due to varibility of realizations of the k comparment-specific concentration for individual i
$e_{M,ijk}$	Part of deviation e_{ijk} due to measurement (assay) error for the kth compartment-specific concentration measured on individual i
$\sigma^2_{R,k}$	Variance associated with variability in realizations of the k comparment-specific concentration within a single individual
$\sigma^2_{M,k}$	Variance associated with measurement (assay) error in the kth compartment-specific concentration
σ^2_k	Total intra-individual variance in concentration measurements on the kth compartment-specific concentration

Figure 1: A very simple PK model: the one-compartment model with first-order absorption and elimination, often used to represent the time course of concentrations of orally-administered drugs.

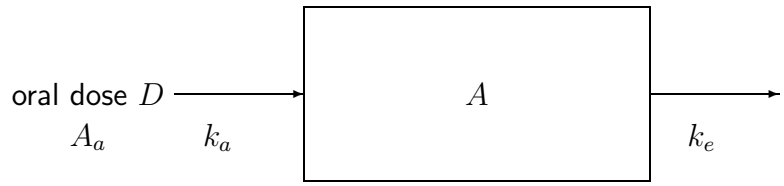

Figure 2: A typical PBPK model with four compartments.

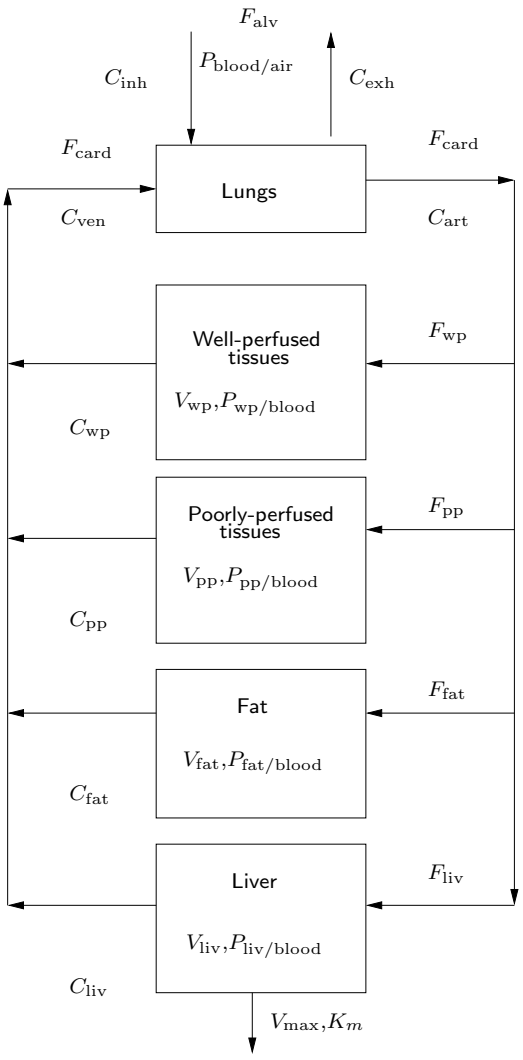

Figure 3: Plasma concentration-time profiles for 12 human subjects receiving the same (scaled to body weight) oral dose of the anti-asthmatic theophylline at time zero.

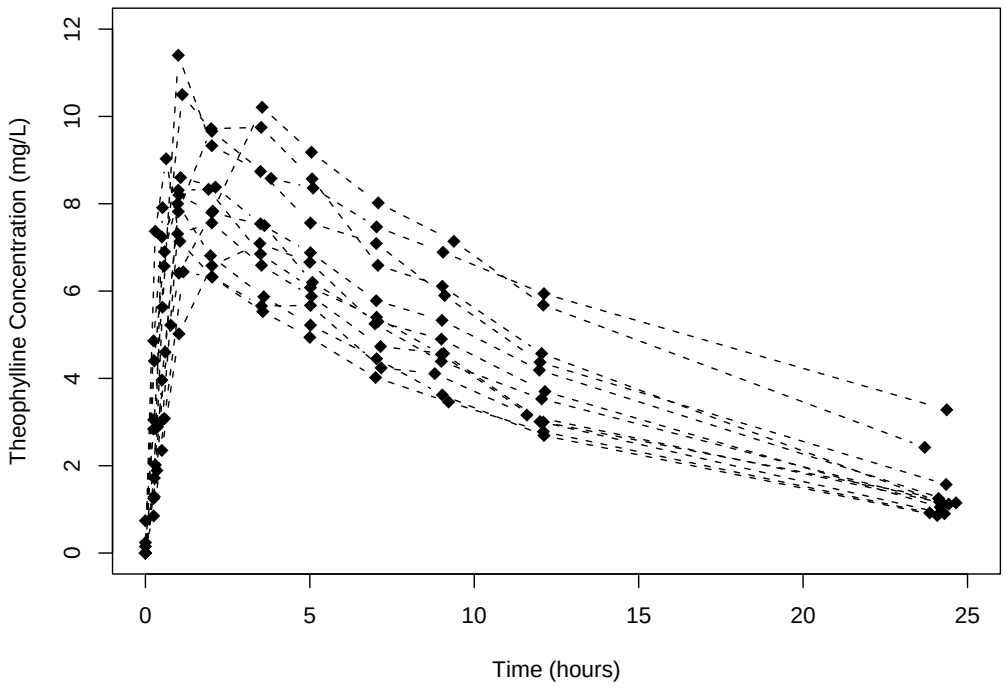

Figure 4: Conceptual depiction of intra-individual sources of variation for exhaled air concentrations. For individual i, the solid black line is $f_k(t, \boldsymbol{E}_i, \boldsymbol{\phi}_i, \boldsymbol{\theta}_i)$, the "inherent trajectory" for i's exhaled air concentrations; the solid gray line is a realization of the true exhaled air concentrations through continuous time; and the solid symbols are intermittent measurements of the realized true concentrations, contaminated by assay measurement error, the data actually observed.